

STIC-Biot ch/ChemLib

Fr m: Ramirez, Delia
Sent: Tuesday, July 09, 2002 6:06 PM
To: STIC-Biotech/ChemLib
Subject: case 09/687230

Hi,

I would like to request the following searches (Brasemann):

1. a standard search of seq id 1 and 2 in the protein databases (commercial and interference)
2. a standard search of seq id 2 151-313 in the protein databases (commercial and interference)

Thank you,

Delia M. Ramirez, Ph.D.
Patent Examiner - Art Unit 1652
USPTO
1911 S. Clark Street, Crystal Mall 1, 10D04, Mail room 10C01
Arlington, VA 22202
(703) 306-0288
delia.ramirez@uspto.gov

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: BOB
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 7-12-02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:17 ; Search time 49.31 Seconds
(without alignments)
2066.397 Million cell updates/sec

Title: US-09-687-230-2
Perfect score: 3073
Sequence: 1 MGKHKHKSHKSHLYEYVE.....PGNMICLLGPSSSEKCLINK 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2990.5	97.3	651	4	Q9NP11
2	2989	97.3	652	4	Q9UH59
3	2744.5	89.3	651	11	Q88665
4	1980	64.4	459	4	Q9BV48
5	1413.5	46.0	351	4	Q96K44
6	699.5	22.8	861	5	Q9VLCX2
7	684	22.3	501	4	Q9H8M2
8	506	16.5	636	5	Q17581
9	413.5	13.5	174	11	Q9CT78
10	391	12.7	233	4	Q9H7R9
11	274	8.9	362	4	Q9H505
12	252.5	8.2	715	4	Q9Y4Q3
13	252	8.2	56	4	Q9UN01
14	231.5	7.5	706	11	Q99JY4
15	229	7.5	805	4	Q95692
16	225.5	7.3	556	10	Q9SFX1

17	222	7.2	652	10	Q9LM88	Q9LM88 arabidopsis
18	218	7.1	1430	5	Q9V4J4	Q9V4J4 drosophila
19	217.5	7.1	952	10	Q9SN19	Q9SN19 oryza sativ
20	208	6.8	2065	5	Q97068	Q97068 drosophila
21	199.5	6.5	1633	13	Q90941	Q90941 gallus gall
22	197	6.4	145	10	Q9F1A2	Q9F1A2 arabidopsis
23	194.5	6.3	1582	4	Q9H301	Q9H301 homo sapien
24	194.5	6.3	1630	13	Q90753	Q90753 gallus gall
25	193.5	6.3	1568	13	Q90755	Q90755 gallus gall
26	192.5	6.3	1235	4	Q9H0E9	Q9H0E9 homo sapien
27	191.5	6.2	578	3	Q9Y7N0	Q9Y7N0 schizosacch
28	191	6.2	729	13	Q90971	Q90971 gallus gall
29	191	6.2	733	13	Q73897	Q73897 gallus gall
30	189.5	6.2	236	11	Q9D007	Q9D007 mus musculu
31	189.5	6.2	1679	4	Q9HBD4	Q9HBD4 homo sapien
32	189	6.2	1582	4	Q9H2T3	Q9H2T3 homo sapien
33	189	6.2	1602	4	Q9H2T5	Q9H2T5 homo sapien
34	189	6.2	1634	4	Q9H314	Q9H314 homo sapien
35	188	6.1	2081	10	Q9LH98	Q9LH98 arabidopsis
36	187.5	6.1	757	4	Q15355	Q15355 homo sapien
37	187	6.1	1647	4	Q9HBD3	Q9HBD3 homo sapien
38	185	6.0	1022	11	Q63928	Q63928 mus sp. brig
39	183	6.0	638	3	Q07442	Q07442 saccharomyc
40	183	6.0	1865	11	Q60544	Q60544 mesocricetu
41	182	5.9	920	4	Q43178	Q43178 homo sapien
42	182	5.9	920	4	Q969M9	Q969M9 homo sapien
43	181.5	5.9	3080	5	Q9VR33	Q9VR33 drosophila
44	181	5.9	798	11	Q54795	Q54795 mus musculu
45	181	5.9	798	11	Q88411	Q88411 mus musculu

ALIGNMENTS

RESULT 1
ID Q9NP11 PRELIMINARY; PRT; 651 AA.
AC Q9NP11:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BROMODOMAIN CONTAINING PROTEIN (NMG4).
GN BPA65.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=PERAL BRAIN;
RC Kziyshkowska J.G., Dobner T.G.;
RA "Cloning of a human bromodomain containing protein."
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yu Y., Li G.Y.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL, AJ271881; CAB72445.1;
DR EMBL, AF152604; AAF75126.1;
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74138 MW; 2987947644C215E7 CRC64;

Query Match 97.3%; Score 2990.5; DB 4; Length 651;
Best local similarity 99.3%; Pred. No. 8.5e-189;
Matches 576; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
1 MGKHKHKSHKSHLYEYVEPKLVKYGNGEYELSTSSGSHDSLSFDKRDHKHKND 60
|||||

```

Db 1 MGKHKHKKSKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEKNDHDKKD 60
QY 61 RRRKKRKKGEQIPIGEEKGRKRRKRVKEDKKRDRVENEAEKDLQCHAPVRLDLPPEKP 120
Db 61 RRRKKRKKGEQIPIGEEKGRKRRKRVKEDKKRDRVENEAEKDLQCHAPVRLDLPPEKP 120
QY 121 LTSSLAKOEVEEQPILOEALNOLMRQLOKRDPSAFSEFPVYDFIAPGYSMITIKHPMFST 180
Db 121 LTSSLAKOEVEEQPILOEALNOLMRQLOKRDPSAFSEFPVYDFIAPGYSMITIKHPMFST 180
QY 181 MKERIKNNNDYOSIELNDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOL 240
Db 181 MKERIKNNNDYOSIELNDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOL 240
QY 241 KOSIDFMAADLOKTRKQKDGDTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
Db 241 KOSIDFMAADLOKTRKQKDGDTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
QY 301 MLEBDFKSNLLEBOEDLDIVKESGKILRRLVNSOCFEERRKPDGTTTGLLHPVDP1 360
Db 301 MLEBDFKSNLLEBOEDLDIVKESGKILRRLVNSOCFEERRKPDGTTTGLLHPVDP1 360
QY 361 VGERGYCLVRLGNTTGRLOSGVNTLQGFKEKDKRNVLPVLYLNGPYSSAAPHYDSTFAN 420
Db 361 VGERGYCLVRLGNTTGRLOSGVNTLQGFKEKDKRNVLPVLYLNGPYSSAAPHYDSTFAN 420
QY 421 ISKDDSDLIYTYGSDSLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIQEME 480
Db 421 ISKDDSDLIYTYGSDSLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIQEME 480
QY 481 MSLEPEDEGHTRLDTGKEMQITVEEPGRLDSDTDRLIALKAVNFGVPEVPESEEA 540
Db 481 MSLEPEDEGHTRLDTGKEMQITVEEPGRLDSDTDRLIALKAVNFGVPEVPESEEA 540
QY 541 EIFOKKIDETTRLLRELOEAONERLSTPRPNNICLLGPS 580
Db 541 EIFOKKIDETTRLLRELOEAONERLSTPRPNNICLLGPS 580

```

```

RESULT 2
ID 09UH59 PRELIMINARY: PRT: 652 AA.
AC 09UH59:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BROMODOMAIN PROTEIN CELTIX1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Steal A., Enserink J., Stein J.L., Stein G.S., van Wijnen A.J.;
RT "Molecular analysis of the human bromodomain protein Celtix-1."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AF213969; AAF19526.1; -.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; Bromodomain.1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO.1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 652 AA; 74084 MW; 70F2B654B2618529 CRC64;

```

```

Query Match 97.3%; Score 2989; DB 4; Length 652;
Best Local Similarity 99.0%; Pred. No. 1.1e-188;
Matches 574; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MGKHKHKKSKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEKNDHDKKD 60
Db 1 MGKHKHKKSKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEKNDHDKKD 60

```

```

QY 61 RRRKKRKKGEQIPIGEEKGRKRRKRVKEDKKRDRVENEAEKDLQCHAPVRLDLPPEKP 120
Db 61 RRRKKRKKGEQIPIGEEKGRKRRKRVKEDKKRDRVENEAEKDLQCHAPVRLDLPPEKP 120
QY 121 LTSSLAKOEVEEQPILOEALNOLMRQLOKRDPSAFSEFPVYDFIAPGYSMITIKHPMFST 180
Db 121 LTSSLAKOEVEEQPILOEALNOLMRQLOKRDPSAFSEFPVYDFIAPGYSMITIKHPMFST 180
QY 181 MKERIKNNNDYOSIELNDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOL 240
Db 181 MKERIKNNNDYOSIELNDNFKLMCTNMIYKPPETIYKKAOKLLHSGMKILSGERTIOL 240
QY 241 KOSIDFMAADLOKTRKQKDGDTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
Db 241 KOSIDFMAADLOKTRKQKDGDTDSGSGEDGCWQEREDSGDAEHAHAKSPSKENKKDKD 300
QY 301 MLEBDFKSNLLEBOEDLDIVKESGKILRRLVNSOCFEERRKPDGTTTGLLHPVDP1 360
Db 301 MLEBDFKSNLLEBOEDLDIVKESGKILRRLVNSOCFEERRKPDGTTTGLLHPVDP1 360
QY 361 VGERGYCLVRLGNTTGRLOSGVNTLQGFKEKDKRNVLPVLYLNGPYSSAAPHYDSTFAN 420
Db 361 VGERGYCLVRLGNTTGRLOSGVNTLQGFKEKDKRNVLPVLYLNGPYSSAAPHYDSTFAN 420
QY 421 ISKDDSDLIYTYGSDSLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIQEME 480
Db 421 ISKDDSDLIYTYGSDSLPDSFSIHEFLATCODYPYVADSLLDVLTKGHSRTIQEME 480
QY 481 MSLEPEDEGHTRLDTGKEMQITVEEPGRLDSDTDRLIALKAVNFGVPEVPESEEA 540
Db 481 MSLEPEDEGHTRLDTGKEMQITVEEPGRLDSDTDRLIALKAVNFGVPEVPESEEA 540
QY 541 EIFOKKIDETTRLLRELOEAONERLSTPRPNNICLLGPS 580
Db 541 EIFOKKIDETTRLLRELOEAONERLSTPRPNNICLLGPS 580

```

```

RESULT 3
ID 08B65 PRELIMINARY: PRT: 651 AA.
AC 08B65:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BROMODOMAIN-CONTAINING PROTEIN BP75.
GN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BALB/C; TISSUE-BRAIN;
RC MEDLINE=99456832; PubMed=10526152;
RA Cuppen E., van Ham M., Peeters B., Wieringa B., Hendriks W.;
RT "Identification and molecular characterization of BP75, a novel
RL bromodomain-containing protein."
EMBL: AF084259; AAC33302.1; -.
DR MGD: MGI:1349766; Brd7.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; Bromodomain.1.
DR SMART: SM00297; BROMO.1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74000 MW; 5D34B4F1AFD51350 CRC64;

```

```

Query Match 89.3%; Score 2744.5; DB 11; Length 651;
Best Local Similarity 88.4%; Pred. No. 1.3e-172;
Matches 518; Conservative 36; Mismatches 31; Indels 1; Gaps 1;

```

```

QY 1 MGKHKHKKSKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEKNDHDKKD 60
Db 1 MGKHKHKKSKHLVEEYERPLKLVKVGNEVELSTGSSGHDLSFEKNDHDKKD 60

```



```

Db 1 MGKKKKKSDRHFEYEEYKPLKLVLYKGSEVTELTSGSGHSDSLFEDRSDDHKND 60
OY 61 RRRKKKKKEKQIPGEKGRKRRRVKEDKKRRDROVENAEKDLOCHAPVRLDLPPEK 120
Db 61 RRRKKKKKEKQAPGEKGRKRRRVKEDKKRRDRAENEDVDLOCHVPIRLDLPPEK 120
OY 121 LRSLSLAKOEEVEPTPLQELNOLMLORQRKDPSEFFSPDTDFIAGYSMIIRHPDFT 180
Db 121 LRSLSLAKOEEVEPTPLQELNOLMLORQRKDPSEFFSPDTDFIAGYSMIIRHPDFT 180
OY 181 MKERIKNNYQSIIEELKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSL 240
Db 181 MKERIKNNYQSIIEELKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSL 240
OY 241 KQSIDPMADLQTRKQKQDGTOSGEGCGMREDESDGAEAHAFKSPKSKKKKDKD 300
Db 241 KQSIDPMADLQTRKQKQDGTOSGEGCGMREDESDGAEAHAFKSPKSKKKKDKD 300
OY 301 MLEDKFKSNLREBOEQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPT 360
Db 301 VLEDKFKSNLREBOEQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPT 360
OY 361 VGEPGYCLVRLGNTGRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFAN 420
Db 361 VGEPGYCLVRLGNTGRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFAN 420
OY 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
Db 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
OY 481 MSLEDEGHTRLDTGKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
Db 481 MSLEDEGHTRLDTGKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
OY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 586
Db 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 586
OY 540 EVFORKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 585
Db 540 EVFORKIDETTRLLRELQEAONERLSTRPPNMICLLGSPSEKIL 585

RESULT 4
OY 4 PRELIMINARY; PRT; 459 AA.
ID 09BVA8
AC 09BVA8;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SIMILAR TO BROMODOMAIN-CONTAINING 7 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN, NEUROBLASTOMA;
RA Strauberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC001611; AA01611.1;
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
FT NON_TER 1
SQ SEQUENCE 459 AA; 51714 MW; 9EF21D1454BCE63B CRC64;

Query Match 64.4%; Score 1980; DB 4; Length 459;
Best Local Similarity 98.7%; Pred. No. 1.7e-122;
Matches 380; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 196 LKDNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSLKQSIDPMADLQTRK 255
Db 3 IODNFKLMCTNAMITYNKPEITYYKAAKLLHSGMKILSQERQSLKQSIDPMADLQTRK 62
OY 256 QNDGDTOSGEGCGMREDESDGAEAHAFKSPKSKKKKDKDLQDKFKSNLREBO 315
Db 256 QNDGDTOSGEGCGMREDESDGAEAHAFKSPKSKKKKDKDLQDKFKSNLREBO 315

```

```

Db 63 QNDGDTOSGEGCGMREDESDGAEAHAFKSPKSKKKKDKDLQDKFKSNLREBO 122
OY 316 EQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPIVGEPGYCLVRLGNT 375
Db 123 EQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPIVGEPGYCLVRLGNT 375
OY 376 GRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFANISKDDSLIYSTYGE 435
Db 183 GRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFANISKDDSLIYSTYGE 435
OY 436 DSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEMENSLPDEGHTRLDT 495
Db 243 DSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEMENSLPDEGHTRLDT 302
OY 496 GKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEAEIFOKKIDETTRLLR 555
Db 303 AKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEAEIFOKKIDETTRLLR 362
OY 556 ELQEAONERLSTRPPNMICLLGPS 580
Db 363 ELQEAONERLSTRPPNMICLLGPS 387

RESULT 5
OY 5 PRELIMINARY; PRT; 351 AA.
ID 096KA4
AC 096KA4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CDNA FLJ144402 FIS, CLONE HEMBA1003783, MODERATELY SIMILAR TO MUS
DE MUSCULUS BROMODOMAIN-CONTAINING PROTEIN BP75 MRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO, AND MAINLY HEAD;
RA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosolri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ichida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawal Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Niomiya K., Iwayanagi T.;
RT "NBD0 human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027308; BAB5031.1;
SQ SEQUENCE 351 AA; 39236 MW; 51E1DD2ABD7DEDA0 CRC64;

Query Match 46.0%; Score 1413.5; DB 4; Length 351;
Best Local Similarity 98.2%; Pred. No. 2.4e-85;
Matches 275; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
OY 301 MLEDKFKSNLREBOEQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPI 360
Db 1 MLEDKFKSNLREBOEQLDRIVKESGGKTLRLVNSQCEFERRRKPDGTTTGLLHPVPI 60
OY 361 VGEPGYCLVRLGNTGRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFAN 420
Db 61 VGEPGYCLVRLGNTGRLSGVNTLQGFKEKDKRNNKVTPLYLNGPYSSYAPHYDSTFAN 120
OY 421 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 480
Db 121 ISKDDSLIYSTYGEDSDLPDSFSIHEFLATCQDYPYVADSLDLVLTGKGSRTLQEME 180
OY 481 MSLEDEGHTRLDTGKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 540
Db 181 MSLEDEGHTRLDTGKEMEOITEVEPPGRILDSSTQDRILAKAVTNFVPEVDSSEA 239
OY 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGPS 580
Db 541 EIFOKKIDETTRLLRELQEAONERLSTRPPNMICLLGPS 580

```

```

DB 240 EIPFKKDETRRLRELOEAONERLSTRPPNMICLLGSPS 279

RESULT 6
09VLMX2 PRELIMINARY; PRT; 861 AA.
ID 09VLMX2: 09VLMX2:
AC 09VLMX2: 09VLMX2:
DT 01-MAR-2001 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG7154 PROTEIN.
OS CG7154.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borovica D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Harris N.L., Harvey D., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hosolt D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington S., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zheng G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003618: AAF52557.1;
DR FlyBase: FBgn0031947; CG7154.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.1.
DR PRINTS: SMO0503; BROMODOMAIN.
DR SMART: SMO0297; BROMO.1.
DR PROSITE: PSS0014; BROMODOMAIN_2.1.
SO SEQUENCE 861 AA; 95922 MW; E27D528E5F9B3AF CRC64;

```

Query Match 22.8%; Score 699.5; DB 5; Length 861;
 Best Local Similarity 27.6%; Pred. No. 7.5e-38;
 Matches 211; Conservative 127; Mismatches 219; Indels 207; Gaps 22;

```

OY 3 KKKKKKSD-KHLYEEY-----VEKPLKLVLYKVGNEVTELTSGSSGD----- 45
DB 5 KKKKKKSSEREKTEEYSQHODPAOLGKLLIKVGSNATPEXSANSPWDGSPPTAAE 64
OY 46 ---SLEFDKNDHDKHKKRRKKRKKGKQIPGEGKRRRVEDKKRRDROVENEAE 102
DB 65 AMSPVPEELQDHGHRHRHKKSKKKKK---KDRKKHKNHNE-KRHRSDHROAGS 119
OY 103 -----KDLQCH----- 108
DB 120 DEDMAGADAAACSGFAPSSVAPPAADPDSSQDSFEMDDQSPLEPILFAGITTDN 179
OY 109 -----APVRLD-----LR----- 116
DB 180 SPSNCPYTKPIAPKLDLILMGSSPNSLSQSSSLGSSPTKPLDILPSTPGGA 239
OY 117 -----PEKPLTSSLAQK-----EVBOTPLQALNOLMROLOKRDPSAF 155
DB 240 NSLNALTPKALEAPKTPSSSSSESGREPRSCVLKLOKSPLNKLEHLRFLKRDPHQF 239
OY 156 FSPVYDFIAPGYSMITHHMDPSTMEKTKNNDYOSIEELKDNFKIMCTNMITNKEPT 215
DB 300 FAMPVYDDMAPGYSTISRPMDSTMRKIDHETALTEFTDFKLMCENAIKYNHVDY 359
OY 216 IYKKAARKLLHSGMKTLISQERI-OSLKOSIDFMADLOKTRKODGTPDSQSGDGCWOR 274
DB 360 VYKKAARKRLQYMKHLQENLRSLKPLSGYREL---TARELGELSSND-----MSR 411
OY 275 ERDSDG-----AEAHAKSPKKNKKDKMDLEKFS-----NNLREQ-BQDOR 320
DB 412 ENNDSADDEGASTGAEEPRPPAOLEEERKRTLLENAPKTHFPVYDDLGLGEILAQOV 471
OY 321 IYKESGKILTRLVNSOCEPERKRPDGTTLGLHHPVDPIVSGCYVRLGTTGRLOS 380
DB 472 AAQQAQVNAKNAKNAKMGFLKMKCTITLNVYEE---NGPFRVYITGLVLRKLOK 528
OY 381 GYNTLOGFEKRNKRVLYLYNYGYSYAPHYDSTFANISKDSLDLYSTGEOSDIP 440
DB 529 GSAOLOTROVDKRNAVRYVKSILNYGAFASFAFDFDSRFSTLSAEIQLVIRTGDASSAE 588
OY 441 SDPSIEFLATODQYVYVADSLDVLTKGHSRTL-----QEMMSLPEDE 487
DB 589 YAESILOFTKQ-SNGTTTANGSLDITNGEKSDELQYNNQMLHSEOREIEKEFEH 647
OY 488 GHTRLDTGKEKEQITVEVPPGRLDSTQDRILALKAVTNFGVYVDFSEAEI----- 542
DB 648 -----ETSSQGETVAQIDQYKKNTHVDPKGLQSLGELGIDVSYFLDGMRAEMKSYEL 701
OY 543 ---FQKKDETRRLRELOEAONERLSTRPPNMICLLGSPSEK 583
DB 702 NRRMEHHSQNLLLEKRLVQAQHDRLS-OPLPNHGLVOPAQOE 744

RESULT 7
09H8M2 PRELIMINARY; PRT; 501 AA.
ID 09H8M2: 09H8M2:
AC 09H8M2: 09H8M2:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CDNA FLJ13441 FIS, CLONE PLAC1002775, WEAKLY SIMILAR TO
DE PREGGRIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Ishigawa T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,

```

```

RESULT      8
017581
ID 017581      PRELIMINARY;      PRT;      636 AA.
AC 017581;
DT 01-NOV-1996 (TEMBREL. 01, Created)
DT 01-NOV-1996 (TEMBREL. 01, Last sequence update)
DT 01-DEC-2001 (TEMBREL. 19, Last annotation update)
DE C01H6.7. PROTEIN.
GN C01H6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae.
OC Rhabditidae; Pelodolerinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
RN (1)
RP SEQUENCE FROM N.A.
RA Berks M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
I21
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; Pubmed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";

```

RESULT	9	
QCST78		
ID	QCST78	PRELIMINARY;
AC	QCST78;	PRT: 174 AA.
DT	01-JUN-2001 (TEMBLrel. 17, Created)	
DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)	
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)	
DE	18 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,	
DE	CLONE:1190001E05, FULL INSERT SEQUENCE (FRAGMENT).	
CS	BBD7.	
OS	Mus musculus (mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;	
FX	MEDLINE=21085660; PubMed=11217651;	
RA	Kawai J., Shindagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.	

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pebole G., Quackenbush J.,
 RA Schraml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Kono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., Fujita M., Gariboldi M.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Yushan-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT Functional annotation of a full-length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 DR EMBL: AK004429; BAB23299.1;
 DR MGD: MGI:1349766; Brd7.
 PT NON_TER 1 1
 SO SEQUENCE 174 AA; 19027 MW; CF47FE3B9D719D03 CRC64;

Query Match 13.5%; Score 413.5; DB 11; Length 174;
 Best Local Similarity 75.2%; Pred. No. 6.3e-20;
 Matches 82; Conservative 11; Mismatches 15; Indels 1; Gaps 1;
 QY 478 EMESLPEDEGHTRTLDGKEMEQITEVEPPGRILSDTOPRLALKNVNGVPEVEFDS 537
 DB 1 DLDMSPDEGGQTRALDPAKEAE-ITQIEPTGRLSSQDRALALQAVTTFGAPAEVFD 59
 QY 538 EEAETOKRKTDETRRLRLRLQEAQNERLSTRPPGNNICLGPSSSEKICL 586
 DB 60 EEAEVFORKLTDETRRLRLRLQEAQNERLSTRPPGNNICLGPSSREXYL 108

RESULT 10
 O9H7R9 PRELIMINARY; PRT; 233 AA.
 AC O9H7R9
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE CDNA FLJ14330 FIS, CLONE PLACE400261, WEAKLY SIMILAR TO PERGRIN.
 DE PERGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y., Oshima A.,
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK024392; BAB14907.1;
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain; 1.
 DR PRINTS: PRO0503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PSS0014; BROMODOMAIN.2; 1.
 SO SEQUENCE 233 AA; 26233 MW; FBF96245BD86F4EB CRC64;

Query Match 12.7%; Score 391; DB 4; Length 233;
 Best Local Similarity 65.1%; Pred. No. 2.8e-18;
 Matches 69; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY 130 EVEDPTIQEALINQMLRQORQPSAFSEFPTDTFLAETSIIKHHPDSTMAKERIKND 189
 DB 19 ENESPTIQQLDLHPTLRQORQDPHGFAPVTDALAPYSMIKHHPDFTMKDKIVANE 78
 QY 190 YOSIEELKDNFKLCTNMIYKPEYIYKAKKLLHSGMKLSOE 235
 DB 79 YKSTFEKADPKLMDNMMYTNRPDIYKAKKLLHSGMKLSOE 124

RESULT 11
 Q9H505 PRELIMINARY; PRT; 362 AA.
 AC Q9H505;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA: FLJ23117 FIS, CLONE LNC10649.
 DE CDNA: FLJ23117 FIS, CLONE LNC10649.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ota T.,
 RA Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y., Ota T., Suzuki Y.,
 RA Ohyashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogal T., Sugano S.,
 RT NEDD human cDNA sequencing project.*;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026830; BAB15565.1;
 SO SEQUENCE 362 AA; 39411 MW; 3B8240D7C70E44A5 CRC64;

Query Match 8.9%; Score 274; DB 4; Length 362;
 Best Local Similarity 25.7%; Pred. No. 2.5e-10;
 Matches 79; Conservative 70; Mismatches 126; Indels 32; Gaps 8;
 QY 291 SKENKKKDKMDLEDFK-----SNNLEFGQQLDRIVKESGKLTIRLVNOCER 340
 DB 31 AKSKKPSSEVISCMEFEPGNACSLTDSAEVIALVSHADADRNRLPGLGKMY 90
 QY 341 ERRKPDGTTTLGLHPDIVGEPGYCLVRLGTTTGRLOSGVTLQGFEDKRNKTPVL 400
 DB 91 LKRNCGSLISVYNTAEPPADDEETHPYDLSLSKLLPGFTTL-GFDERNNVT--- 146
 QY 401 YNTGPGSSYARHYDSTFANISKDSLDLYTYGEDSDLPSPSHETLATQDPYVNA 460
 DB 147 FLS-SATVIALSMQNNVSFEDLKSDEWELLYSAYGDETVQCALSIQEFVKDAGSYSKV 205
 QY 461 DSLDLVLTGKHSRTLOEKE-----MSLPEDEGHTRTL-DGKMEQITEVEPPGRIDS 513
 DB 206 DLLDLPITGDSRILFLQKORNNPKRPDPAKVGDTIGDSSSYLEMENK----- 258
 QY 514 STQDRILAKAVTNGVPEVDFSEALITPQKLTDETRRLRLRLQEAQNERLSTRPPGNN 573
 DB 239 STPDVSVISLMSLSGKVKKEIDPDSHL---NLDETTRKLQDLHQAQERNGSGSPSSL 315
 QY 574 ICLLGP 580
 DB 316 SLSNMS 322
 RESULT 12
 Q9Y4Q3 PRELIMINARY; PRT; 715 AA.
 AC Q9Y4Q3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 79.7 KDA PROTEIN (FRAGMENT).


```

DN      01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE      D5J24E15.1 (PERGRIN (Br140 PROTEIN)) (FRAGMENT)..
GN      D5J24E15.1.
OS      Homo sapiens (Human) .
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Palmer S.;
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: Z84485; CAB06488.1; -.
DR      InterPro: IPR001487; Bromodomaiin.
DR      InterPro: IPR001965; PHD.
DR      InterPro: IPR000313; PWMP.
DR      Pfam: PF00439; bromodomaiin: 1.
DR      Pfam: PF00628; PHD: 1.
DR      Pfam: PF00855; PWMP: 1.
DR      PRINTS: PRO0503; BROMODOMAIN.
DR      SMART: SM00297; BROMO: 1.
DR      SMART: SM00249; PHD: 2.
DR      SMART: SM00293; PWMP: 1.
DR      PROSITE: PS50014; BROMODOMAIN_2; 2.
FT      NON_TER      1      805      1
SQ      SEQUENCE      805 AA; 90651 MW; E28C017F5C545334 CRC64;

```

Query Match

Query Match	7.5%;	Score 229;	DB 4;	Length 805;
Best Local Similarity	26.1%;	Pred. No. 6.4e-07;		
Matches 67;	Conservative 55;	Mismatches 107;	Indels 28;	Gaps 6

[illegible]

Search completed: July 11, 2002, 15:56:16
Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:49:34 ; Search time 55.8 Seconds
(without alignments)
324,463 Million cell updates/sec

Title: us-09-687-230-2_COPY_151_313

Perfect score: 85
Sequence: 1 DP5AFSFPYDTFAPGYSM.....NKKKDKMDLEDKFSNNIER 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855	100.0	589	19	AAW37947 Phosphatidylinositol
2	855	100.0	667	22	AAO13504 Human polypeptide
3	855	100.0	718	22	AAU16626 Human novel secret
4	375	43.9	501	22	AA895421 Human protein sequ
5	356	41.6	405	22	AAW3712 Human polypeptide,
6	328	38.4	233	22	AA895881 Human protein sequ
7	328	38.4	280	21	AA841780 Human ORF1544
8	328	38.4	597	22	AAW58835 Human polypeptide
9	328	38.4	605	22	AAW40621 Human polypeptide
10	284.5	33.3	861	22	AB863028 Drosophila melanog
11	174	20.4	414	22	AA863767 Human prostate can

12	174	20.4	715	22	ABG12634 Novel human diago
13	174	20.4	1058	22	AAW39231 Human polypeptide
14	174	20.4	1109	22	AAW41017 Human polypeptide
15	174	20.4	1109	22	AAW41018 Human polypeptide
16	174	20.4	1189	22	AAW39232 Human polypeptide
17	169.5	19.8	616	21	AAW38052 Arabidopsis thalia
18	169	19.8	707	22	ABG21300 Novel human diago
19	163	19.1	1173	22	ABG21301 Novel human diago
20	163	19.1	2543	22	ABG21295 Novel human diago
21	160	18.7	1430	22	ABW58602 Human prostate can
22	155	18.1	270	22	AAW38879 Human prostate can
23	153	17.9	245	22	AAW38828 Novel human diago
24	151	17.7	1924	22	ABG06006 Drosophila melanog
25	147	17.2	513	22	ABW68341 Novel human secret
26	147	17.2	969	22	AAW30116 Protein regulating
27	146	17.1	951	21	AAW58634 Drosophila melanog
28	146	17.1	2065	22	ABW66356 Novel human diago
29	145.5	17.0	346	22	ABW60006 TATA-binding prote
30	145.5	17.0	1872	17	AAW60078 TATA-binding prote
31	145.5	17.0	1872	17	AAW60078 TATA-binding prote
32	145.5	17.0	1872	18	AAW25030 Novel human diago
33	145.5	17.0	1886	22	ABW60009 TATA-binding prote
34	145.5	17.0	1893	15	AAW56491 Human TATA-binding
35	145.5	17.0	1893	17	AAW60082 Human TATA-binding
36	145.5	17.0	1893	18	AAW25020 Human tumour suppr
37	144	16.8	1646	21	AAW27553 Human tumour suppr
38	144	16.8	1647	21	AAW27549 Human tumour suppr
39	144	16.8	1649	21	AAW27557 Human tumour suppr
40	144	16.8	1650	21	AAW27555 Human tumour suppr
41	144	16.8	1654	22	ABW56684 Drosophila melanog
42	144	16.8	1678	21	AAW27554 Human tumour suppr
43	144	16.8	1679	21	AAW27552 Human tumour suppr
44	144	16.8	1681	21	AAW27558 Human tumour suppr
45	144	16.8	1682	21	AAW27536 Human tumour suppr

ALIGNMENTS

RESULT 1	AAW37947	standard; Protein: 589 AA.
ID	AAW37947	
XX	AAW37947	
AC	AAW37947	
XX	14-SEP-1998	(first entry)
DT	14-SEP-1998	
XX		
DE	Phosphatidylinositol-3' kinase associated protein.	
XX		
KW	Phosphatidylinositol-3' kinase associated protein; PI3K; PIKAP;	
KW	human; signal transduction; cell growth; cancer; restenosis;	
KW	therapy; diagnosis.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	151..313
FT	Region	/note="Bromodomain"
FT		516..589
FT		/note="PI3K p85 binding region"
PN	W09820126-A1.	
XX		
PD	14-MAY-1998.	
XX		
PP	01-OCT-1997;	97WO-US15845.
XX		
PR	01-NOV-1996;	96US-0030103.
XX		
PA	(ONXX-) ONXX PHARM INC.	
XX		
PI	Brasellmann S;	
XX		

DR WPI: 1998-286942/25.
DR N-PSDB: AAV29267.
XX
PT New isolated phosphatidylinositol-3' kinase associated protein -
XX used to develop products for diagnosis and treatment of cell growth
XX disorders such as restenosis or cancer
PS Claim 10; Page 40-41; 52pp; English.

CC This polypeptide comprises human phosphatidylinositol-3' kinase
CC (PI3K) associated protein (PIKAP), a protein that binds to the
CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
CC which exhibits a bromodomain. Its amino acid sequence was deduced
CC from a cDNA clone (see AAV29267) obtained from a HeLa library using
CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
CC provides vectors containing nucleic acid sequences that encode
CC PIKAP or its fragments, host cells, methods for the expression of
CC PIKAP, and methods for using the products for the diagnosis and
CC treatment of cell growth disorders such as restenosis or cancer.
CC Also described is an assay for identifying agonists and antagonists
CC of PI3K regulation. These include mutant PIKAPs that compete with
CC native PIKAPs for binding to PI3K, antibodies, and nucleotide
CC sequences that can be used to inhibit or enhance PIKAP gene
CC expression. Transgenic and knock-out animals are also described.

CC Sequence 589 AA:

Query Match 100.0%; Score 855; DB 19; Length 589;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DPSAFSPFVTDFIAPGYSMIIRKHPMDFSMKEIKNNQYSEELKDNFKLCTNMTY 60
DB 151 dpaaffsfvtdfiapgyssmllkhpmdfscmkelkndyqsllelkdnlklnctnamty 210
OY 61 NKRETIYYKAARKLLHSGMKILSOERISQKSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 211 nkpetiyykaarkllhsgmkilsgerlqslgsidfmadlqtrkqkdgtqsgsedg 270
OY 121 CMQREEDSGDAHAHAFKSPSKENKKKDDMLDKFSNNLER 163
DB 271 cwgrededsgdaaahafkspskenkdkdmlcdkfknnler 313

RESULT 2

AA013504
ID AA013504 standard; Protein; 667 AA.

AC AA013504;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 27396.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PE 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR N-PSDB: AAI93435.

PT Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -

PS Claim 20; SEQ ID NO 27396; 1399pp + Sequence listing; English.

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

CC Sequence 667 AA:

Query Match 100.0%; Score 855; DB 22; Length 667;
Best Local Similarity 100.0%; Pred. No. 1.4e-74;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DPSAFSPFVTDFIAPGYSMIIRKHPMDFSMKEIKNNQYSEELKDNFKLCTNMTY 60
DB 166 dpaaffsfvtdfiapgyssmllkhpmdfscmkelkndyqsllelkdnlklnctnamty 225
OY 61 NKRETIYYKAARKLLHSGMKILSOERISQKSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 226 nkpetiyykaarkllhsgmkilsgerlqslgsidfmadlqtrkqkdgtqsgsedg 285
OY 121 CMQREEDSGDAHAHAFKSPSKENKKKDDMLDKFSNNLER 163
DB 286 cwgrededsgdaaahafkspskenkdkdmlcdkfknnler 328

RESULT 3

AAU16626
ID AAU16626 standard; Protein; 718 AA.

AC AAU16626;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 1579.

XX Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; optalmalogical;
KW vulnery; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; anglogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

PR 17-JAN-2001; 2001WO-US01341.

XX (HYSE-) HYSEQ INC.

PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

Claim 11: SEQ ID No 1579; 980pp: English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention.

Query Match 100.0%; Score 855; DB 22; Length 718;
 Best Local Similarity 100.0%; Pred. No. 1.5e-74;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSAFFSFPVTDPIAGYSMIIRKHPDFTMKERIKNDYOSIELKDNFKLMCTNAMY 60
 DB 218 dpaaffsfvtdpiagysmiikhpmdfstmketkndyosielkdnfkmlctnamiy 277
 QY 61 NKPEITYYKAARKLHSGKMTLSQERIOSLQSIDFMADLOKTRKDKGDTQSQSEDDG 120
 DB 278 nkpetiyykaarklhsgkmtlsqerioslqsidfmadlqtrkdkgdtdsqseddg 337
 QY 121 CMQREHEDSGDAVAHAFKSPKSKENKKDDMLDKFKSNLNR 163
 DB 338 cmqreedsdgdaahafkspskenkkddmledfkfksnlner 380

RESULT 4
 AAB95421
 ID AAB95421 standard; Protein: 501 AA.

AC AAB95421;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:17823.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

XX

PT primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

PS

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 8: SEQ ID 17823; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 501 AA:

Query Match

Best Local Similarity 43.9%; Score 375; DB 22; Length 501;

Matches 77; Conservative 28; Mismatches 37; Indels 20; Gaps 3;

QY 1 DSAFFSFPVTDPIAGYSMIIRKHPDFTMKERIKNDYOSIELKDNFKLMCTNAMY 60
 DB 40 dphgffaifvtdaiagysmiikhpmdfstmketkndyosielkdnfkmlctnamiy 99
 QY 61 NKPEITYYKAARKLHSGKMTLSQERIOSLQSIDFMADLOKTRKDKGDTQSQSEDDG 120
 DB 100 nrpdlvyykklakklhagfkmskerllakrsmfmgdm-----dfsqgaallg 149
 QY 121 CMQREHEDSGDAVAHAFKSPKSKENKKDDMLDKFKSNLNR 157
 DB 150 -----nedlaveepvpyvvetakskkprrevlscmte 186

RESULT 5

AAM93712
 ID AAM93712 standard; Protein: 405 AA.

AC AAM93712;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3652.

KW Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

XX

PD 05-SEP-2001.
 XX 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94661.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS
 PS Claim 8; SEQ ID NO 3652; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 405 AA;
 Query Match 41.6%; Score 356; DB 22; Length 405;
 Best Local Similarity 100.0%; Pred. No. 2.5e-26;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 MADLQKTRKQKQDPTSGSGGSGCWORERSDGDAFAHAFKSPSKKKKKDKMLQDKF 156
 Db 1 madlqktrkqkqdcstsgsgsgcworeredsgdaeahakspkknkkdkmlqdkf 60
 QY 157 KSNMLER 163
 Db 61 ksnmler 67
 RESULT 6
 AAB95881
 ID AAB95881 standard; Protein: 233 AA.
 AC AAB95881;
 XX
 XX 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:18979.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 PS Claim 8; SEQ ID 18979; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 XX
 SQ Sequence 233 AA;
 Query Match 38.4%; Score 328; DB 22; Length 233;
 Best Local Similarity 67.1%; Pred. No. 6.5e-24;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;
 QY 1 DPSAFSPFVTDIAPGYSMLIKHPMDFTMKKEIKNNQYOSIEBLKDNFKLMTNAMY 60
 Db 40 dphgfafprvdalagysmltkhpmdfgtmkdkivaneysvtefkedfklmcdnamty 99
 QY 61 NKPEITYYKAARKLIHSGMKILSOE 85
 Db 100 nrpdtvyyklakklhagfkmmksk 124
 RESULT 7
 AAB41780
 ID AAB41780 standard; Protein: 280 AA.
 AC AAB41780;
 XX
 XX 08-FEB-2001 (first entry)
 DE Human ORFX ORF1544 polypeptide sequence SEQ ID NO:3088.
 XX
 XX Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;
 KW vulnerability; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thromolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypertoid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 XX
 PN WO200058473-A2.
 PD
 PD 05-OCT-2000.
 XX
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 DR
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC75989.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS
 PS Claim 11; Page 2306-2307; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparasitism; nootropic; neuroprotective;
 CC osteoactive; anticonvulsant; antiallergic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antihistaminic; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy. The
 CC vectors. The proteins and nucleic acids may be used to treat cancers;
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SQ Sequence 280 AA:

Query Match 38.4%; Score 328; DB 21; Length 280;
 Best Local Similarity 67.1%; Pred. No. 8 2e-24;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;
 OY 1 DSAFSPFVDTFAPGSMITKHPDSTMKETKNNDYOSIEELKDNFKIMCTNAMY 60
 DB 81 ophgftatpvtalapysmilkpmdfgtmkdklvaneykvtelkadfkimcdnamy 140
 OY 61 NKPEITYYKAARKLHSGMKILSOE 85
 DB 141 nrdptlvvykklakklhagfkmksq 165
 RESULT 8

AAM38835
 ID AAM38835 standard; Protein; 597 AA.
 XX
 XX
 AC AAM38835;
 XX
 XX
 DT 22-OCT-2001 (first entry)
 XX
 XX
 DE Human polypeptide SEQ ID NO 1980.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 PD
 PD 26-JUL-2001.
 XX
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-NOV-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI57991.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Example 3; SEQ ID NO 1980; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC activation/inhibition of the activities such as: immune system suppression,
 CC and thrombolytic activity, chemotactic/chemokinetic activity, haemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SQ Sequence 597 AA:

Query Match 38.4%; Score 328; DB 22; Length 597;
 Best Local Similarity 67.1%; Pred. No. 2 1e-23;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;
 OY 1 DSAFSPFVDTFAPGSMITKHPDSTMKETKNNDYOSIEELKDNFKIMCTNAMY 60
 DB 141 nrdptlvvykklakklhagfkmksq 165

Db 295 dphqfawpvdmdnapyslirpmdfstmrqkiddheylalefiddfkImcealky 354
QY 61 NKPEITYYKAARKLLHSGMKTLISOERI-QSLKOSIDFMADIQKROKDGNDTSQSGHDG 119
Db 355 nhvdytnkaakrlllygmkhlgpenlmrslklsygmrel---tateIgfelsnd--- 408
QY 120 GCMQREHEDSGD-----AEAHAFKSPSKENKKDKMDLEDKFSN 159
Db 409 --msrendadagastgtaeeptrpaqlleeerktlrIlenapth 452

RESULT 11

AAB63767
ID AAB63767 standard; Protein; 414 AA.

AC AAB63767;

DT 26-MAR-2001 (first entry)

DE Human prostate cancer associated antigen protein sequence SEQ ID NO.1129.

KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDWIG) LUDWIG INST CANCER RES.

PI Obata Y;

DR WPI; 2001-025274/03.

PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.,
PT cancer -

PS Example 1; Page 708-709; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (CAAP) respectively.
XX AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63870
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterized by expression of an abnormal amount of a protein,
XX e.g. cancer.

CC Sequence 414 AA;

Query Match 20.4%; Score 174; DB 22; Length 414;
Best Local Similarity 38.4%; Pred. No. 1.3e-08;

Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY 1 DSAFSPFVPTDFIAPGYSMTIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMIY 60
Db 134 dparifagvslkevpdyldhikhpmdfatmrkrlaegyknlhefeedfdllidncmy 193

QY 61 NKPEITYYKAARKLLHSGMKTLISOER 86

Db 194 nardtlfyaaavrlrtdggsvvlrpar 219

RESULT 12

ABG12634
ID ABG12634 standard; Protein; 715 AA.

AC ABG12634;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #12625.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS76821.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 42993; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (II) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.

CC Sequence 715 AA;

Query Match 20.4%; Score 174; DB 22; Length 715;
Best Local Similarity 38.4%; Pred. No. 2.5e-08;

Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

QY 1 DSAFSPFVPTDFIAPGYSMTIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMIY 60
Db 108 dparifagvslkevpdyldhikhpmdfatmrkrlaegyknlhefeedfdllidncmy 167

QY 61 NKPEITYYKAKKLLHSGMKILSOER 86
 Db 168 nardtlyfyaavrlrdggvvlrqar 193

RESULT 13

AAM39231
 ID AAM39231 standard; Protein: 1058 AA.

XX AAM39231;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2376.

XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

DR N-PSDB; AAI58387.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Example 4; SEQ ID NO 2376; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 1058 AA;

Query Match 20.4%; Score 174; DB 22; Length 1058;
 Best Local Similarity 38.4%; Pred. No. 4.2e-08;
 Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0;

RESULT 14

AAM41017
 ID AAM41017 standard; Protein: 1109 AA.

XX AAM41017;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 5948.

XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

DR N-PSDB; AAI60173.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 5948; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:09; Search time 22.7 Seconds

(without alignments)
175.391 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DSAFFSPVDTFIAPGYSM.....NKKKDKMLEDKFKSNLNR 163

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	855	100.0	589	4	US-08-942-008-2
2	145.5	17.0	1872	1	US-08-188-582-14
3	145.5	17.0	1872	1	US-08-646-715-14
4	145.5	17.0	1893	1	US-08-188-582-11
5	145.5	17.0	1893	1	US-08-646-715-11
6	129	15.1	2441	3	US-08-194-468-2
7	129	15.1	2441	3	US-08-961-739-2
8	123	14.4	65	1	US-08-227-536-5
9	123	14.4	65	5	PCT-US95-04682-5
10	122	14.3	2414	1	US-08-227-536-2
11	122	14.3	2414	5	PCT-US95-04682-2
12	118	13.8	65	1	US-08-227-536-6
13	118	13.8	65	5	PCT-US95-04682-6
14	108	12.6	65	1	US-08-227-536-7
15	108	12.6	65	5	PCT-US95-04682-7
16	103	12.0	65	1	US-08-227-536-8
17	103	12.0	65	5	PCT-US95-04682-8
18	98	11.5	65	1	US-08-227-536-4
19	98	11.5	65	5	PCT-US95-04682-4
20	94	11.0	65	1	US-08-227-536-3
21	94	11.0	65	5	PCT-US95-04682-3
22	91.5	10.7	976	4	US-08-894-997-50
23	80	9.4	562	2	US-08-973-675-2
24	79.5	9.3	914	1	US-08-487-890A-11
25	79.5	9.3	914	2	US-08-478-435-11
26	79.5	9.3	914	2	US-08-337-483-11
27	79.5	9.3	914	2	US-08-478-373-11

28	79.5	9.3	914	3	US-08-474-671-11	Sequence 11, Appl
29	79.5	9.3	914	3	US-08-483-577A-11	Sequence 11, Appl
30	79.5	9.3	914	4	US-08-897-428-11	Sequence 11, Appl
31	79.5	9.3	914	4	US-08-637-654-11	Sequence 11, Appl
32	79	9.2	676	2	US-08-398-590A-40	Sequence 40, Appl
33	79	9.2	676	4	US-08-894-997-40	Sequence 40, Appl
34	77.5	9.1	2482	1	US-08-328-254-6	Sequence 4, Appl
35	77.5	9.1	3248	1	US-08-353-700-1	Sequence 1, Appl
36	77.5	9.1	3248	5	PCT-US95-16216-1	Sequence 20, Appl
37	76.5	8.9	356	4	US-09-308-003-20	Sequence 1, Appl
38	76	8.9	1618	1	US-07-853-913-4	Sequence 4, Appl
39	75.5	8.8	258	2	US-08-602-941-3	Sequence 3, Appl
40	75.5	8.8	258	3	US-08-961-264-3	Sequence 3, Appl
41	75	8.8	608	2	US-08-736-770-1	Sequence 1, Appl
42	74.5	8.7	576	2	US-08-533-306A-2	Sequence 2, Appl
43	74.5	8.7	576	2	US-08-742-923A-2	Sequence 2, Appl
44	74.5	8.7	816	2	US-08-533-306A-6	Sequence 6, Appl
45	74.5	8.7	816	2	US-08-742-923A-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-08-942-008-2
Sequence 2, Application US/08942008
Patent No. 6133419
GENERAL INFORMATION:
APPLICANT: Basellmann, Sylvia
TITLE OF INVENTION: Nucleotide Sequences that Encode
TITLE OF INVENTION: Phosphatidylinositol-3' Kinase Associated Proteins and
TITLE OF INVENTION: Uses Thereof
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESS: ONYX Pharmaceuticals, Inc.
STREET: 3031 Research Drive
CITY: Richmond
STATE: CA
COUNTRY: USA
ZIP: 94806
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,008
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Giotta, Gregory
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX1027
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 262-8710
TELEFAX: (510) 222-9758
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-008-2

Query Match 100.0%; Score 855; DB 4; Length 589;
Best Local Similarity 100.0%; Pred. No. 2, 6e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSAFFSPVDTFIAPGYSMIIKHPMDFSTMKERIKNNQYSGIEELKDNFKIMCTNAMTY 60
DB 151 DSAFFSPVDTFIAPGYSMIIKHPMDFSTMKERIKNNQYSGIEELKDNFKIMCTNAMTY 210
```

QY 61 NRPEITYYKAKKLLHSGMKILISOERIQSLKOSIDFMADLOKTRKOKGDTGTSQSGEDG 120
DB 211 NRPEITYYKAKKLLHSGMKILISOERIQSLKOSIDFMADLOKTRKOKGDTGTSQSGEDG 270
QY 121 CWOEREDSGDAFAHAFKSPSKENKKDKMDLDRKSNLTER 163
DB 271 CWOEREDSGDAFAHAFKSPSKENKKDKMDLDRKSNLTER 313

RESULT 2
US-08-188-582-14; Sequence 14, Application US/08188582
; Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dylact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarradero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-188-582-14

Query Match 17.0%; Score 145.5; DB 1; Length 1872;

Best Local Similarity 26.0%; Pred. No. 6.3e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPVYDFLAPGYSMTIKHPMDSFSTMEKIKNDYOSIEELKDNFKLMCNAMITKPPET 65
DB 1528 FHPVKKKFPDYKYIVNPMDELIRKNISKKHKYOSRESFLDDVNLILANSVKYGPES 1587
QY 66 IYKAAKLLHSGMKILSQ--ERISLK-----QSIDFMADLOKTRKOK 107
DB 1588 QYTKIAOEIVNCYQTLTFDEHLTQLEKDICTAKAALAEALFESLDPMTPGPYTPQPP 1647
QY 108 DGTDTQS---GDDGCGWQRRERDS--GDAEAHAFKSPSKENKKDKMDLEDR 155

DB 1648 DLYDTNLSMSRSDASVFODESMSVLDIPSATPEKQVTOEGEDGGLADEE 1700

RESULT 3
US-08-646-715-14; Sequence 14, Application US/08646715
; Patent No. 5637686

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dylact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TATS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarradero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-646-715-14

Query Match 17.0%; Score 145.5; DB 1; Length 1872;

Best Local Similarity 26.0%; Pred. No. 6.3e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPVYDFLAPGYSMTIKHPMDSFSTMEKIKNDYOSIEELKDNFKLMCNAMITKPPET 65
DB 1528 FHPVKKKFPDYKYIVNPMDELIRKNISKKHKYOSRESFLDDVNLILANSVKYGPES 1587
QY 66 IYKAAKLLHSGMKILSQ--ERISLK-----QSIDFMADLOKTRKOK 107
DB 1588 QYTKIAOEIVNCYQTLTFDEHLTQLEKDICTAKAALAEALFESLDPMTPGPYTPQPP 1647
QY 108 DGTDTQS---GDDGCGWQRRERDS--GDAEAHAFKSPSKENKKDKMDLEDR 155
DB 1648 DLYDTNLSMSRSDASVFODESMSVLDIPSATPEKQVTOEGEDGGLADEE 1700

RESULT 4
US-08-188-582-11
Sequence 11, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1893 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-11

Query Match 17.0%; Score 145.5; DB 1; Length 1893;
Best Local Similarity 26.0%; Pred. No. 6.4e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPPTDFTIAGYSMTIIRHMPDFTMKRIRKNDYQSIIEELKDNFKLCTNAMIYKPEP 65
DB 1549 FHHPRNKKFVVDYKYVINPMDLETIRKNIKSKYQRESFLDYNLLIANSVKYNGPES 1608
QY 66 IYKAKKLHSGMKILSQ--ERIOSLK-----OSIDFMADLQKTRQK 107
DB 1609 QYTKAQLVAVCYQTLFEYDEHLTQLEKDICTAKEALEAELESLDPMFPGPYTPQPP 1668
QY 108 DGTDTSS--GEDGCMQREKEDS--GDAEHAHAKSPSKENKKDKMDLEK 155
DB 1669 DLYDTNTSLMSRDSVFDQESNMSVLDIPSATPEKQVTOEGEDGDDGLADEE 1721

RESULT 5
US-08-646-715-11
Sequence 11, Application US/08646715
Patent No. 5637686
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio

APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1893 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-11

Query Match 17.0%; Score 145.5; DB 1; Length 1893;
Best Local Similarity 26.0%; Pred. No. 6.4e-07;
Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSPPTDFTIAGYSMTIIRHMPDFTMKRIRKNDYQSIIEELKDNFKLCTNAMIYKPEP 65
DB 1549 FHHPRNKKFVVDYKYVINPMDLETIRKNIKSKYQRESFLDYNLLIANSVKYNGPES 1608
QY 66 IYKAKKLHSGMKILSQ--ERIOSLK-----OSIDFMADLQKTRQK 107
DB 1609 QYTKAQLVAVCYQTLFEYDEHLTQLEKDICTAKEALEAELESLDPMFPGPYTPQPP 1668
QY 108 DGTDTSS--GEDGCMQREKEDS--GDAEHAHAKSPSKENKKDKMDLEK 155
DB 1669 DLYDTNTSLMSRDSVFDQESNMSVLDIPSATPEKQVTOEGEDGDDGLADEE 1721

RESULT 6
US-08-194-468-2
Sequence 2, Application US/08194468
Patent No. 5750336
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESS: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPILED: IBM PC DOS 3.31

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0

CURRENT APPLICATION DATA: W 1.0, VERSION #1.23

APPLICATION NUMBER: US/08/194,468

FILING DATE: 10-FEB-1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFO: BMA

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,

REFERENCE/DOCKET NUMBER: P41 9672

ELECOMMUNICATION INFORMATION:
TEL EDUHOV: 6510-5545 FAX: 4737

TELEPHONE: (619)-546-473
TELEFAX: (619)-546-0303

TELEFAX: (619)-546-9392

SEQUENCE CHARACTERISTICS:

LENGTH: 2441 amino acids

TYPE: amino acid

TOPOLOGY: linear

OLECULE TYPE: protein

700 E. C.T.

Query Match	15.18;	Score 129;	DB 1;	Length 2441;
Best Local Similarity	27.08;	Pred. No. 5.8e-05;		
Matches	51;	Conservative	25;	Mismatches 69;
				Indels 44;
				Gaps 9

QY	1	DPSAF-ESEPTDID--APGSMIKIRHMFSTWKEKIKINNDYQSTEEELKNEFKMLCTNA	57
Db	1106	DPESPEFQOPDPOLLGIPYFEDIVKNNPMDLSTIKRKLDITGQYQEPWQYVDVRLMFNA	1166
QY	58	MIYKRPETIYVKAARKILHSGMKILISOERIQSLKOSIDEMADIDQTKRKROKOGTR-----	112
Db	1166	WLYNRKTSRYVKFECSK-----AEYFEDE-IDPYMOSLGYCCG---RKKEFSPPQLCCYG	1216
QY	113	-----SQSGEDGCGW-----QREREDSGDAEHAHAFKSPKKNKKRMDME	153
Db	1217	KQLCTIPDAAYYSYQNRVHFCCGCKPTEIIGENAVITLGD-----DPSPQOTTISKDPE	1266
QY	154	DKFSNNMLE 162	
Db	1270	KK-KNDTLD 1277	

```

RESULT 7
US-08-961-739-2
: Sequence 2, Application US/08961739A
: Patient No. 6063583
: GENERAL INFORMATION:
: APPLICANT: Montanuy, Marc R.
: TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
: FILE REFERENCE: SALK1650-1
: CURRENT APPLICATION NUMBER: US/08/961,739A
: CURRENT FILING DATE: 1997-10-31
: EARLIER APPLICATION NUMBER: US 194,468
: EARLIER FILING DATE: 1994-02-10
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 2441
: TYPE: PRT
: ORGANISM: Mus
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(2441)

```

OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

Query Match	15.1%;	Score 129;	DB 3;	Length 2441;
Best Local Similarity	27.0%;	Pred. No. 5.8e-05;		
Matches	51;	Conservative 25;	Mismatches 69;	Indels 44;
			Gaps	9

[illegible]

```

      RESULT      8
US-08-227-536-5
: Sequence 5, Application US/08227536
: Patent No. 5658784
:
: GENERAL INFORMATION:
: APPLICANT: Eckner, Richard
: APPLICANT: Ewen, Mark
: APPLICANT: Livingston, David
: TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
: TITLE OF INVENTION: FACTOR P300 AND USBS OF P300
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
: STREET: Ten Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02109
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/227,536
: FILING DATE: 14-APR-1994
: CLASSIFICATION: 436
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams P.D., Kathleen A.
: REGISTRATION NUMBER: 34,380
: REFERENCE/DOCKET NUMBER: DFCI-308XX
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-2290
: TELEFAX: (617) 451-0313
: INFORMATION FOR SRO ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 65 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
:
US-08-227-536-5
Query Match      14.4%      Score 123; DB 1; Length 65
Best Local Similarity 47.9%      Pred. No. 1,6e-06;

```

Matches 23; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

QY 16 PGSMIIKHPDMFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 63
 18 PDYHKIHKQPMDMGTIKRLENNYIWAASECMQDFNTMTNCTIYKNP 65

RESULT 9
 PCT-US95-04682-5
 Sequence 5, Application PC/TUS9504682

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/227,536

FILING DATE: 14-April-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Holliday C. Heine, Ph.D.

REGISTRATION NUMBER: 34,346

REFERENCE/DOCKET NUMBER: DECI-308X999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 65 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

PCT-US95-04682-5

Query Match 14.4%; Score 123; DB 5; Length 65;

Best Local Similarity 47.9%; Pred. No. 1.6e-06;

Matches 23; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

QY 16 PGYSMIKHPDMFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 63
 18 PDYHKIHKQPMDMGTIKRLENNYIWAASECMQDFNTMTNCTIYKNP 65

RESULT 10

US-08-227-536-2

Sequence 2, Application US/08227536

Patent No. 5658784

GENERAL INFORMATION:

APPLICANT: Eckner, Richard

APPLICANT: Even, Mark

APPLICANT: Livingston, David

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/227,536

FILING DATE: 14-Apr-1994

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Williams Ph.D., Kathleen A.

REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: DECI-308XX

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2414 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-227-536-2

Query Match 14.3%; Score 122; DB 1; Length 2414;

Best Local Similarity 34.3%; Pred. No. 0.00033;

Matches 34; Conservative 15; Mismatches 42; Indels 8; Gaps 4;

QY 1 DPGAF-ESFPVYDFI--APGYSMIKHPDMFSTMKERIKNDYOSIELKDNFKLMCTNMIYKNP 57
 1069 DPESLPFQPVDPOLGIDPFDIVKSPMDLSTIRKIDTGOYQBPWOYVDIVIMFNNA 1128

QY 58 MIYNKPEITYYKAAKILHSGMKILSOBRISLKSIDF 96
 DB 1129 WLYNKRKTSRVYKSKL-----SEVEQF-IDPVMSLGI 1162

RESULT 11

PCT-US95-04682-2

Sequence 2, Application PC/TUS9504682

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

TITLE OF INVENTION: FACTOR P300 AND USES OF P300

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/227,536

FILING DATE: 14-April-1994

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	14.3%	Score 122;	DB 5;	Length 2414;
	Best Local Similarity	34.3%;	Pred. NO. 0.00033;		
	Matches	34;	Conservative	15;	Mismatches 42;
				Indels	8; Gaps
OY	1 DPSAF-FSPPTWDFI--APGYSMIRKHMDSFTMEKIKNNDIYSIEELKDNFKLACTNA	57			
	: :	: : : : : : :	:	:	:
Db	1069 DPESLPEFQPVDPOLLGIPDIFDYVKSMDISTIRKLDTCGOYQEPWOYVDIDIMFNNA	1128			
	: :	: : : : : : :	:	:	:
OY	58 MLYNKPETIYYAAKKILHSGAKIIISGRIOISKOSIDE	96			
	::: : :	: : : : : : :	:	:	:
Db	1129 WLYNRKTSRYVKYCKSL-----SEYFEQL-IDPVMOSTLG Y	1162			
	::: : :	: : : : : : :	:	:	:

RESULT 12
 US-08-227-536-6
 ; Sequence 6, Application US/08227536
 ; Patent No. 5658784
 ;
 GENERAL INFORMATION:
 ;
 APPLICANT: Eckner, Richard
 APPLICANT: Ewen, Mark
 APPLICANT: Livingston, David
 TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
 TITLE OF INVENTION: FACTOR P300 AND USES OF P300
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: US
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/227, 536
 FILING DATE: 14-APR-1994
 CLASSIFICATION: 436
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams Ph.D., Kathleen A.
 REGISTRATION NUMBER: 34,380
 REFERENCE/DOCKET NUMBER: DPC1-308XX
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 452-2290
 TELEFAX: (617) 451-0313
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 65 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal

```

US-08-227-536-6
Query Match      13.8%; Score 118; DB 1; Length 65;
Best Local Similarity 47.8%; Pred. No. 5.8e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 0; Gaps 0;
QY 18 YSMIRKHPDFTSMKERTKNDYOSIEELKDKQFKMTCTNMAIYNKP 63
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 YHDIIRKHPMDLSIVYRKRMKNNDYRKAQCPAAVLRMTFSCNYKYNKP 65

```

```

1  RESULT 13
2  PCT-US95-04682-6
3  ; Sequence 6, Application PC/TUS9504682
4  ; GENERAL INFORMATION:
5  ; APPLICANT:
6  ; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
7  ; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
8  ; NUMBER OF SEQUENCES: 13
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: Weingarten, Schurigin, Gagnebin & Hayes
11 ; STREET: Ten Post Office Square
12 ; CITY: Boston
13 ; STATE: MA
14 ; COUNTRY: US
15 ; ZIP: 02109
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Floppy disk
18 ; COMPUTER: IBM PC compatible
19 ; OPERATING SYSTEM: PC-DOS/MS-DOS
20 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: PCT/US95/04682
23 ; FILING DATE:
24 ; CLASSIFICATION:
25 ; PRIOR APPLICATION DATA:
26 ; APPLICATION NUMBER: US 08/227,536
27 ; FILING DATE: 14-April-1994
28 ; CLASSIFICATION:
29 ; ATTORNEY/AGENT INFORMATION:
30 ; NAME: Holliday C. Heine, Ph.D.
31 ; REGISTRATION NUMBER: 34,346
32 ; REFERENCE/DOCKET NUMBER: DFCI-308x999
33 ; TELECOMMUNICATION INFORMATION:
34 ; TELEPHONE: (617) 542-2290
35 ; TELEFAX: (617) 451-0313
36 ; INFORMATION FOR SEQ ID NO: 6:
37 ; SEQUENCE CHARACTERISTICS:
38 ; LENGTH: 65 amino acids
39 ; TYPE: amino acid
40 ; TOPOLOGY: linear
41 ; MOLECULE TYPE: peptide
42 ; HYPOTHETICAL: NO
43 ; ANTI-SENSE: NO
44 ; FRAGMENT TYPE: internal
45 ;
46 PCT-US95-04682-6

```

```

Query Match      13.8%; Score 118; DB 5; Length 65;
Best Local Similarity 47.8%; Pred. No. 5.8e-06;
Matches 22; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY      18 YSMIIKHPMDSTYMKETIKNDYOSIEELKDNFKIMCTNAMIKNP 63
      | | | | | | | | | | : : : | | : : | | | | |
Db      20 YHDIIKHPMDSTYVKRMENRDYRDAGEADYRLMFSCNYKNPF 65

RESULT 14
US-08-227-536-7
; Sequence 7, Application US/08227536
; Patent No. 5658784
GENERAL INFORMATION:

```

APPLICANT: Eckner, Richard
APPLICANT: Even, Mark
APPLICANT: Livingston, David
TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,536
FILING DATE: 14-APR-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Williams Ph.D., Kathleen A.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: DFCI-308XX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-227-536-7

Query Match 12.6%; Score 108; DB 1; Length 65;
Best Local Similarity 36.0%; Pred. No. 7.2e-05;
Matches 18; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

QY 12 DFLAPGYSMIRKHPDEFSTMKRIRKNDYQSIETELKDNFKLMCTNMIYN 61
DB 14 ELXHPYELIRKPVDFKRIKIRIRNHKYSIGDLEKDVMLCHNAQTFN 63

RESULT 15
PCT-US95-04682-7
Sequence 7, Application PC/TUS9504682
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
TITLE OF INVENTION: FACTOR P300 AND USES OF P300
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04682
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/227,536
FILING DATE: 14-April-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Holliday C. Heine, Ph.D.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-308Xg999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
PCT-US95-04682-7

Query Match 12.6%; Score 108; DB 5; Length 65;
Best Local Similarity 36.0%; Pred. No. 7.2e-05;
Matches 18; Conservative 15; Mismatches 17; Indels 0; Gaps 0;

QY 12 DFLAPGYSMIRKHPDEFSTMKRIRKNDYQSIETELKDNFKLMCTNMIYN 61
DB 14 ELXHPYELIRKPVDFKRIKIRIRNHKYSIGDLEKDVMLCHNAQTFN 63

Search completed: July 11, 2002, 15:50:11
Job time: 549 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:50:43 ; Search time 29.85 Seconds
(without alignments)
524.709 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DPSAFSEFPVTDFIAPGYSM.....NKKKDKMLDEKKSNNLER 163

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	200.5	23.5	636	2 T18845	hypothetical prote
2	174	20.4	715	2 T12534	hypothetical prote
3	169	19.8	1214	2 JC2069	zinc-finger protei
4	163.5	19.1	556	2 D96791	hypothetical prote
5	158.5	18.5	1572	2 S45251	SNF2alpha protei
6	158.5	18.5	1586	2 S39580	HRM protein - hum
7	151.5	17.7	757	2 S68142	probable transcrip
8	148.5	17.4	454	2 T37933	transcription acti
9	146.5	17.1	733	2 T28145	RING3 kinase - chl
10	146	17.1	2068	2 A47371	transcription init
11	145.5	17.0	1893	1 A40262	transcription init
12	144	16.8	1022	2 T53078	homeotic gene regu
13	144	16.8	1613	2 S39059	protein BRC1 - hum
14	144	16.8	1647	2 S45252	SNF2beta protei
15	142	16.6	754	2 A56619	female sterile hom
16	140	16.4	374	2 T33328	hypothetical prote
17	135.5	15.8	1865	1 A43742	transcription init
18	135	15.8	2038	2 T40984	female sterile hom
19	133.5	15.6	578	2 T40984	transcription fact
20	132	15.4	1450	2 S32373	DNA-binding protei
21	131.5	15.4	832	1 S71788	F/CAP protein - hu
22	130.5	15.3	638	2 S67605	hypothetical prote
23	130	15.2	439	1 S28051	transcription fact
24	129	15.1	2441	2 S39161	CREB-binding prote
25	128	15.0	1633	2 JC5056	polyoma 1 - chic
26	124.5	14.6	1680	2 T41628	probable transcrip
27	124	14.5	1638	2 A42091	transcription acti
28	123	14.4	2440	2 S39162	transcription coac
29	122	14.3	1879	2 T19481	hypothetical prote

30	122	14.3	2414	2 A54277	transcription adap
31	120	14.0	542	2 S54260	bromodomain protei
32	118	13.8	449	2 T12495	hypothetical prote
33	117	13.7	1332	2 S41552	probable transcrip
34	115.5	13.5	979	2 T40006	hypothetical prote
35	115	13.5	1479	2 T17401	transcription regu
36	114	13.3	361	2 T42517	bromodomain protei
37	112	13.1	400	2 T00472	probable RING3 pro
38	111.5	13.0	369	2 T46098	hypothetical prote
39	111	13.0	86	2 T47620	histon acetyltrans
40	110.5	12.9	1474	2 T20488	hypothetical prote
41	110	12.9	1251	2 B86194	hypothetical prote
42	110	12.9	2027	2 S60123	hypothetical prote
43	110	12.9	2056	2 G88564	protein R10E1.1
44	105.5	12.3	1250	2 T22845	hypothetical prote
45	105.5	12.3	1703	2 S15047	SNF2 protein - yea

ALIGNMENTS

RESULT 1

hypothetical protein C01H6.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000

C:Accession: T18845

R:Bierks, M.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z19030

A:Accession: T18845

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-636 <M1>

A:Cross-references: EMBL:Z19030; PIDN:CAA95779.1; GSPDB:GN00019; CESP:C01H6.7

A:Experimental source: clone C01H6

C:Genetics:

A:Gene: CESP:C01H6.7

A:Map position: 1

A:Insertions: 20/1; 198/2; 265/3; 451/3; 489/3; 525/3

C:Superfamily: bromodomain homology

F:172-227/Domain: bromodomain homology <BRO>

Query Match 23.5%; Score 200.5; DB 2: Length 636;

Best Local Similarity 31.6%; Pred. No. 1.3e-08; Matches 54; Conservative 32; Mismatches 56; Indels 29; Gaps 5;

QY 1 DPSAFSEFPVTDFIAPGYSMIIRKHPMDFSTMKERIKRNDYOSIEELKDNFKIMCTNAMY 60

DB 167 DPEGYFAFPVTPSKAPYRDIITKTPMDLQITRENIENIGKIASLPAMKEDCELIYSNAFY 226

QY 61 NKEETIYKAKKLLHSGMKILSGERIOSLSKSIDF-----NADLOKTR-- 104

DB 227 NOPWTVVYLAARKLSNLIAVYFGQYLRFLPHSLPMANKIPFELVIGIRLAPVPERKRTMN 286

QY 105 KQMDGTTSQSGDEGCGWQ-----REREDSGDAEAFAPSPSENKKKK 149

DB 287 KRAAVRVKDMGTSD--CLQVADPKVRERLSAKLPEAN-----NPKKKMK 330

RESULT 2

hypothetical protein DKFZp434B094.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000

C:Accession: T12534

R:Wambutt, R.; Heuener, D.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z17524

A:Accession: T12534

A>Status: preliminary

A:Molecule type: mRNA

Nature 408, 816-820, 2000

HBRM protein - human

[illegible]

A:Residues: 1372-1379;1386-1604, 'DNE','CSSKANDVICLIQYSSQIEELRF' <SE6>
A:Cross-references: EMBL:X07024
R:Ruppert, S.; Wang, E.H.; Tjian, R.
Nature 362, 175-179, 1993
A>Title: Cloning and expression of human TAF(II)250: a TBP-associated factor implicated in transcription initiation
A:Reference number: S32352
A:Accession: S32352
A:Molecule type: mRNA
A:Residues: 178-198 <RUP>
R:Hisatsake, K.; Hasegawa, S.; Takada, R.; Nakatani, Y.; Horikoshi, M.; Roeder, R.G.
Nature 362, 179-181, 1993
A>Title: The p250 subunit of native TATA box-binding factor TFIIID is the cell-cycle regulated protein p34^{cdc2}
A:Reference number: S32353; MUID:93196705
A:Accession: S32353
A>Status: preliminary
A:Molecule type: protein
A:Residues: P;1587-595;1009-1022;1351-1355;1357-1360 <HIS>
A>Note: 1351-val, 1353-Lys, and 1354-Glu were also found
C:Genetics:
A:Gene: GDB:TAF2A; CCG1; CCGS; NSCL2; TAFII250
A:Cross-references: GDB:120573; OMIM:313650
A:Map position: Xq13.1-Xq13.1
C:Superfamily: transcription initiation factor IID 250k chain; bromodomain homology;
C:Keywords: alternative splicing; cell cycle control; DNA binding; duplication; phosphatase
F:1-1893/Product: transcription Initiation factor IID 250k chain splice form I status
F:1-177-199-1893/Product: transcription Initiation factor IID 250k chain splice form F
F:1216-1295/Domains: HMGB box homology <HMGI>
F:1372-1379/Region: nuclear location signal
F:1426-1481/Domains: nuclear location signal
F:1549-1604/Domains: bromodomain homology <BRD2>
F:1737-1740/1751-1847/1871/Binding site: phosphate (Ser) (covalent) (by casein kinase F:678,1054,1684/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase F:1020,1361/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) \$state F:1381,1400/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase)

THIS PAGE BLANK (USPTO)


```

DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50812; PWM; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KM Zinc-finger; Bromodomain.
FT ZN_FING 21 47 C2H2-TYPE.
FT ZN_FING 273 323 PHD-TYPE.
FT ZN_FING 386 400 C4-TYPE.
FT DOMAIN 645 715 BROMODOMAIN.
FT DOMAIN 1085 1168 PWM.
FT CONFLICT 299 299 E -> L (IN REF. 2).
FT CONFLICT 729 729 V -> L (IN REF. 2).
SQ SEQUENCE 1214 AA; 137542 MW; C530CD2FE3083A53D CRC64;

Query Match 19.8%; Score 169; DB 1; Length 1214;
Best Local Similarity 37.5%; Pred. No. 2,7e-06;
Matches 36; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 1 DPSAFSPVPTDFIAPGYSMIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMY 60
DB 648 DTGNFSEPVLPSEVPLDHIKKPMDFPMKQNLAVRYLNFDDFEEDFNLYSNCLKY 707
OY 61 NKPEITYKAKKILHSGMKILSGERIOSLQSIDF 96
DB 708 NAKDTIFRAVRLREGGAVRQARRQAEKMGIDF 743

RESULT 4
ID SN22_HUMAN STANDARD; PRT; 1586 AA.
AC PS1531;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Possible global transcription activator SNF2L2 (SNF2-alpha).
GN SNARCA2 OR SNF2L2 OR BRM OR SNF2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=94038910; PubMed=8223438;
RA Muchardt C., Yaniv M.;
RT "A human homologue of Saccharomyces cerevisiae SNF2/SMI2 and
RT Drosophila brm genes potentiates transcriptional activation by the
RT glucocorticoid receptor.";
RL EMBL J. 12:4279-4290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=94268902; PubMed=8208605;
RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brhma are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor.";
RL Nucleic Acids Res. 22:1815-1820(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72889; CAAS1407.1; -
DR EMBL: D26155; BAA05142.1; -
DR MIM: 600014; -
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00439; Bromodomain_1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase; Alternative splicing.
FT DOMAIN 216 238 POLY-GLN.
FT DOMAIN 241 249 POLY-GLN.
FT DOMAIN 555 558 POLY-ARG.
FT DOMAIN 639 646 POLY-GLU.
FT SITE 745 752 ATP (POTENTIAL).
FT NP_BIND 847 850 DEGR BOX.
FT DOMAIN 1293 1297 POLY-GLU.
FT DOMAIN 1415 1485 BROMODOMAIN.
FT DOMAIN 1514 1525 POLY-GLU.
FT VARSPIC 1397 1414 MISSING (IN SHORT ISOFORM).
FT CONFLICT 239 239 P -> POOP (IN REF. 2).
FT CONFLICT 390 390 Q -> E (IN REF. 2).
FT CONFLICT 509 509 G -> S (IN REF. 2).
FT CONFLICT 707 707 W -> R (IN REF. 2).
FT CONFLICT 1135 1135 C -> H (IN REF. 2).
FT CONFLICT 1390 1390 D -> V (IN REF. 2).
SQ SEQUENCE 1586 AA; 180762 MW; FA537E2A2392807A CRC64;

Query Match 18.5%; Score 158.5; DB 1; Length 1586;
Best Local Similarity 30.3%; Pred. No. 2,7e-05;
Matches 43; Conservative 27; Mismatches 47; Indels 25; Gaps 4;

QY 16 PGYSMIKHPMDFSTMKERIKNDYOSIEELKDNFKIMCTNAMYKPEITYYAAKLL 75
DB 1433 PEYELINKPVPDFPKIRKINRHRYSLGDEKVMMLCHNAQGFNLEGSQYEDSYL- 1491
OY 76 HSGMKILSGERIOSLQSIDFMADLQTRKOKDGTDSQSGEDGCMQREHEDSGDAEAH 135
DB 1492 -----OSV-FKSAROKTAKEESEDSNEE-----EEDEESESSEAK 1529
QY 136 AFKSPSKENKKRDKMLEDKR 157
DB 1530 SVKVKIKLKKRDKG--RDKGK 1549

RESULT 5
ID T2D1_DROME STANDARD; PRT; 2068 AA.
AC PS1123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIID 230 kDa subunit (TAFII-230)
DE (TAFII250) (TBP-associated factor 230 kDa) (P230).
GN TAF250.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 63-75 AND 540-546.

```


DR PROSITE: PS50014; BROMODOMAIN_2; 2.
 KW Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
 KM Transcription regulation; Phosphorylation.
 FT DOMAIN 157 165 PRO-RICH.
 FT DNA_BIND 1195 1273 HMG BOX (POTENTIAL).
 FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1397 1467 BROMODOMAIN 1.
 FT DOMAIN 1520 1590 BROMODOMAIN 2.
 FT DOMAIN 1627 1872 ASF/GLU-RICH (ACIDIC TAIL).
 SQ SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64.

Query Match 17.0%; Score 145.5; DB 1; Length 1872;
 Best Local Similarity 26.0%; Pred. No. 0.00038;
 Matches 45; Conservative 33; Mismatches 72; Indels 23; Gaps 4;

QY 6 FSEFVTEIAPGYSMIRKHPMDEFTMEKIKNDYQSEIEELKDNFKIMCTNAMYKRPET 65
 DB 1528 FHHPVNNKKFVDPYKIVNPNDELTIKRNKSKHQSRSLDVLNLSVKNPES 1587
 QY 66 IYVAAKKLHSGMKILSQ--ERIQSEK-----QSIDFMADLQTRKOK 107
 DB 1588 QYTKAEIYVWCYQTLTEYDEHLTQLEKDICTAKEALEAELESIDPMTPGYTPQPP 1647
 QY 108 DGTDTSS---GEGGCGMQREREDS--GDAFAHAFKSPSKENKKDKMDLEDK 155
 DB 1648 DLYDTNLSMSRDASVAFQDESMSVLDIPSPATPEKOVYTOGEGDGDLEDE 1700

RESULT 7
 SN24_HUMAN STANDARD; PRT; 1647 AA.
 AC P5132;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Possible global transcription activator SNF2L4 (SNF2-beta) (BRG-1 protein) (Mitotic growth and transcription activator) (Brahma protein homolog 1).
 GN SMARCA4 OR SNF2L4 OR BRG1 OR SNF2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCBL_TaxID=9606.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94050144; PubMed=8232556;
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RT "BRG1 contains a conserved domain of the SM12/SNF2 family necessary for normal mitotic growth and transcription.";
 RL Nature 366:170-174(1993).
 RN [2]
 RP REVISIONS.
 RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94268902; PubMed=8208605;
 RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
 RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brhma are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor.";
 RL Nucleic Acids Res. 22:1815-1820(1994).
 RN [4]
 RP SEQUENCE OF 814-1474 FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V., Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J., Dangnan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T., Trankhelm M., Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;

RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR HORMONE RECEPTOR TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D29175; AAB40977.1; -
 DR EMBL; D26156; BAA05143.1; -
 DR EMBL; AC006127; AAC97987.1; -
 DR MIM; 603254; -
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00176; SNF2_N; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC_C; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KM ATP-binding; Helicase.
 FT DOMAIN 578 588 POLY-LYS.
 FT DOMAIN 663 672 POLY-GLU.
 FT NP_BIND 779 786 ATP (POTENTIAL).
 FT SITE 881 884 DEB BOX.
 FT DOMAIN 1360 1364 POLY-GLU.
 FT DOMAIN 1477 1547 BROMODOMAIN.
 FT DOMAIN 1571 1584 POLY-GLU.
 SQ SEQUENCE 1647 AA; 184585 MW; 7B785E7953277F1D CRC64.

Query Match 16.8%; Score 144; DB 1; Length 1647;
 Best Local Similarity 27.1%; Pred. No. 0.00044;
 Matches 39; Conservative 31; Mismatches 44; Indels 30; Gaps 5;

QY 16 PGYMIIRKHPMDEFTMEKIKNDYQSEIEELKDNFKIMCTNAMYKRPETIYKAAKKL 75
 DB 1495 PEYELIRKPYDFKIKIRIRNHRSLNDEKVMILCONAQTFNDEGSLEYEDS---- 1550
 QY 76 HSGMKILISOERIOSKISIPFADLQTRKOKDGTDSQSGEDGCGMQREREDSGDAEAH 135
 DB 1551 ----IYQSYFTYVRKIE-----KDD-----SEGEES-----EEEGEGEEGS 1586
 QY 136 AFKSPSKENKKK--DKDLEDFK 157
 DB 1587 ESESRSVVKIKIKRKEKAQDRK 1610

RESULT 8
 BRD3_HUMAN STANDARD; PRT; 726 AA.
 ID BRD3_HUMAN
 AC Q15059; Q92645;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bromodomain-containing protein 3 (RING3-like protein).
 GN BRD3 OR RING3L OR KIA00043.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=9232974; PubMed=1352711;
RA Beck S., Hanson I., Kelly A., Pappin D.J.C., Trowsdale J.;
RT "A homologue of the Drosophila female sterile homeotic (fsh) gene in
RL the class II region of the human MHC.";
RN DNA Seq. 2:203-210(1992).
RN [2]
RP REVISIONS TO N-TERMINUS.
RX MEDLINE=96376536; PubMed=8781126;
RA Thorpe K.L., Abdulla S., Kaufman J., Trowsdale J., Beck S.;
RT "Phylogeny and structure of the RING3 gene.";
RN Immunogenetics 44:391-396(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RX Sato S., Nagase T., Seki T., Ishikawa K.-I., Tabata S.;
RT Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
RN -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-slb.ch).
CC -----
DR EMBL; X62083; CAA4396.1; -;
DR EMBL; M80613; AAG8890.1; ALT_INIT.
DR EMBL; D42040; BAA07641.1; -;
DR PIR; S18860; S18860.
DR PIR; S40781; S40781.
DR MIM; 601340; -;
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain_2.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SMO0297; BROMO_2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR Bromodomain; Repeat; Nuclear protein.
RW Bromodomain 1 163
FT FT 1 364 436 BROMODOMAIN 1.
FT FT 2 476 515 BROMODOMAIN 2.
FT FT 3 546 566 GLU/SER-RICH.
FT FT 4 566 575 ARG/LYS-RICH (HIGHLY BASIC).
FT FT 5 575 801 SER-RICH.
FT FT 6 801 859 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT 7 859 861 ET DOMAIN.
FT FT 8 861 881 POLY-PRO.
FT FT 9 881 906 POLY-GLU.
FT FT 10 906 951 POLY-LYS.
FT FT 11 951 959 POLY-GLU.
FT FT 12 959 964 POLY-GLU.
FT FT 13 964 975 POLY-SER.
FT FT 14 975 993 POLY-SER.
FT FT 15 993 1000 POLY-SER.
FT FT 16 1000 1000 POLY-SER.
FT FT 17 1000 1000 POLY-SER.
FT FT 18 1000 1000 POLY-SER.
FT FT 19 1000 1000 POLY-SER.
FT FT 20 1000 1000 POLY-SER.
FT FT 21 1000 1000 POLY-SER.
FT FT 22 1000 1000 POLY-SER.
FT FT 23 1000 1000 POLY-SER.
FT FT 24 1000 1000 POLY-SER.
FT FT 25 1000 1000 POLY-SER.
FT FT 26 1000 1000 POLY-SER.
FT FT 27 1000 1000 POLY-SER.
FT FT 28 1000 1000 POLY-SER.
FT FT 29 1000 1000 POLY-SER.
FT FT 30 1000 1000 POLY-SER.
FT FT 31 1000 1000 POLY-SER.
FT FT 32 1000 1000 POLY-SER.
FT FT 33 1000 1000 POLY-SER.
FT FT 34 1000 1000 POLY-SER.
FT FT 35 1000 1000 POLY-SER.
FT FT 36 1000 1000 POLY-SER.
FT FT 37 1000 1000 POLY-SER.
FT FT 38 1000 1000 POLY-SER.
FT FT 39 1000 1000 POLY-SER.
FT FT 40 1000 1000 POLY-SER.
FT FT 41 1000 1000 POLY-SER.
FT FT 42 1000 1000 POLY-SER.
FT FT 43 1000 1000 POLY-SER.
FT FT 44 1000 1000 POLY-SER.
FT FT 45 1000 1000 POLY-SER.
FT FT 46 1000 1000 POLY-SER.
FT FT 47 1000 1000 POLY-SER.
FT FT 48 1000 1000 POLY-SER.
FT FT 49 1000 1000 POLY-SER.
FT FT 50 1000 1000 POLY-SER.
FT FT 51 1000 1000 POLY-SER.
FT FT 52 1000 1000 POLY-SER.
FT FT 53 1000 1000 POLY-SER.
FT FT 54 1000 1000 POLY-SER.
FT FT 55 1000 1000 POLY-SER.
FT FT 56 1000 1000 POLY-SER.
FT FT 57 1000 1000 POLY-SER.
FT FT 58 1000 1000 POLY-SER.
FT FT 59 1000 1000 POLY-SER.
FT FT 60 1000 1000 POLY-SER.
FT FT 61 1000 1000 POLY-SER.
FT FT 62 1000 1000 POLY-SER.
FT FT 63 1000 1000 POLY-SER.
FT FT 64 1000 1000 POLY-SER.
FT FT 65 1000 1000 POLY-SER.
FT FT 66 1000 1000 POLY-SER.
FT FT 67 1000 1000 POLY-SER.
FT FT 68 1000 1000 POLY-SER.
FT FT 69 1000 1000 POLY-SER.
FT FT 70 1000 1000 POLY-SER.
FT FT 71 1000 1000 POLY-SER.
FT FT 72 1000 1000 POLY-SER.
FT FT 73 1000 1000 POLY-SER.
FT FT 74 1000 1000 POLY-SER.
FT FT 75 1000 1000 POLY-SER.
FT FT 76 1000 1000 POLY-SER.
FT FT 77 1000 1000 POLY-SER.
FT FT 78 1000 1000 POLY-SER.
FT FT 79 1000 1000 POLY-SER.
FT FT 80 1000 1000 POLY-SER.
FT FT 81 1000 1000 POLY-SER.
FT FT 82 1000 1000 POLY-SER.
FT FT 83 1000 1000 POLY-SER.
FT FT 84 1000 1000 POLY-SER.
FT FT 85 1000 1000 POLY-SER.
FT FT 86 1000 1000 POLY-SER.
FT FT 87 1000 1000 POLY-SER.
FT FT 88 1000 1000 POLY-SER.
FT FT 89 1000 1000 POLY-SER.
FT FT 90 1000 1000 POLY-SER.
FT FT 91 1000 1000 POLY-SER.
FT FT 92 1000 1000 POLY-SER.
FT FT 93 1000 1000 POLY-SER.
FT FT 94 1000 1000 POLY-SER.
FT FT 95 1000 1000 POLY-SER.
FT FT 96 1000 1000 POLY-SER.
FT FT 97 1000 1000 POLY-SER.
FT FT 98 1000 1000 POLY-SER.
FT FT 99 1000 1000 POLY-SER.
FT FT 100 1000 1000 POLY-SER.
FT FT 101 1000 1000 POLY-SER.
FT FT 102 1000 1000 POLY-SER.
FT FT 103 1000 1000 POLY-SER.
FT FT 104 1000 1000 POLY-SER.
FT FT 105 1000 1000 POLY-SER.
FT FT 106 1000 1000 POLY-SER.
FT FT 107 1000 1000 POLY-SER.
FT FT 108 1000 1000 POLY-SER.
FT FT 109 1000 1000 POLY-SER.
FT FT 110 1000 1000 POLY-SER.
FT FT 111 1000 1000 POLY-SER.
FT FT 112 1000 1000 POLY-SER.
FT FT 113 1000 1000 POLY-SER.
FT FT 114 1000 1000 POLY-SER.
FT FT 115 1000 1000 POLY-SER.
FT FT 116 1000 1000 POLY-SER.
FT FT 117 1000 1000 POLY-SER.
FT FT 118 1000 1000 POLY-SER.
FT FT 119 1000 1000 POLY-SER.
FT FT 120 1000 1000 POLY-SER.
FT FT 121 1000 1000 POLY-SER.
FT FT 122 1000 1000 POLY-SER.
FT FT 123 1000 1000 POLY-SER.
FT FT 124 1000 1000 POLY-SER.
FT FT 125 1000 1000 POLY-SER.
FT FT 126 1000 1000 POLY-SER.
FT FT 127 1000 1000 POLY-SER.
FT FT 128 1000 1000 POLY-SER.
FT FT 129 1000 1000 POLY-SER.
FT FT 130 1000 1000 POLY-SER.
FT FT 131 1000 1000 POLY-SER.
FT FT 132 1000 1000 POLY-SER.
FT FT 133 1000 1000 POLY-SER.
FT FT 134 1000 1000 POLY-SER.
FT FT 135 1000 1000 POLY-SER.
FT FT 136 1000 1000 POLY-SER.
FT FT 137 1000 1000 POLY-SER.
FT FT 138 1000 1000 POLY-SER.
FT FT 139 1000 1000 POLY-SER.
FT FT 140 1000 1000 POLY-SER.
FT FT 141 1000 1000 POLY-SER.
FT FT 142 1000 1000 POLY-SER.
FT FT 143 1000 1000 POLY-SER.
FT FT 144 1000 1000 POLY-SER.
FT FT 145 1000 1000 POLY-SER.
FT FT 146 1000 1000 POLY-SER.
FT FT 147 1000 1000 POLY-SER.
FT FT 148 1000 1000 POLY-SER.
FT FT 149 1000 1000 POLY-SER.
FT FT 150 1000 1000 POLY-SER.
FT FT 151 1000 1000 POLY-SER.
FT FT 152 1000 1000 POLY-SER.
FT FT 153 1000 1000 POLY-SER.
FT FT 154 1000 1000 POLY-SER.
FT FT 155 1000 1000 POLY-SER
```

QY 123 QREDSGDAAHAF-----KSPSKENKKK 149
 Db 506 ESESDSEERAHRLAEOLRAVHEOLALSGPISKPKRKKKKKKKK 559

RESULT 10
 FSH_DROME STANDARD; PRT: 2038 AA.
 ID FSH_DROME
 AC P13709; P13710; 13, Created
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Female sterile homeotic protein (Fragile-chorion membrane protein).
 GN FSH1H OR FSH.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=69276730; PubMed=2567251;
 RA Haynes S.R., Mazer B.A., Bhatia-Dey N., Dawid I.B.;
 RT "The Drosophila fsh locus, a maternal effect homeotic gene, encodes
 RT apparent membrane proteins."
 RL Dev. Biol. 134:246-257(1989).
 CC -1- FUNCTION: REQUIRED MATERIALLY FOR PROPER EXPRESSION OF OTHER
 CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
 CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23221; AAA28540.1; -;
 DR EMBL: M23222; AAA28541.1; ALT_TERM.
 DR EMBL: M15762; AAA70424.1; -;
 DR EMBL: M15763; AAA70423.1; -;
 DR EMBL: M15764; AAA70422.1; -;
 DR PIR: A43742; A43742.
 DR HSP: P04002; 1WFA.
 DR FlyBase: FBgn0004656; fs(1)h.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 2.
 DR PRINTS: PRO0503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 2.
 DR PROSITE: PS00633; BROMODOMAIN_1; 2.
 DR PROSITE: PS50014; BROMODOMAIN_2; 2.
 KW Developmental Protein; Bromodomain; Transmembrane; Repeat.
 FT DOMAIN 51 123 BROMODOMAIN 1.
 FT DOMAIN 51 123 BROMODOMAIN 2.
 FT DOMAIN 495 567 ET DOMAIN.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 FT TRANSMEM 750 770 POTENTIAL.
 FT TRANSMEM 790 810 POTENTIAL.
 FT TRANSMEM 816 830 POTENTIAL.
 FT TRANSMEM 874 894 POTENTIAL.
 FT TRANSMEM 1731 1751 POTENTIAL.
 FT TRANSMEM 1939 1959 POTENTIAL.
 FT VARIANT 909 909 G -> A.
 FT VARIANT 1022 1022 H -> RPY.
 SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

Query Match 15.88; Score 135; DB 1; Length 2038;
 Best Local Similarity 28.44; Pred. No. 0.0031;

Matches 42; Conservative 22; Mismatches 70; Indels 14; Gaps 5;
 QY 6 FSPFYDFI-----APGSMITHPMDFSTMKRKNNDQSTIEELKDNFKLCTNAMI 59
 Db 55 FSWPFOQPVDAKKLLNLPDYHKLTIKQPMDSIKRLENNYWSAKETIQDNTFNHCYV 114
 QY 60 YNKEETIYKRAKKLLSGMKILSQERIQSL-KQSIDFMADLQTRKQDGTDSGSD 118
 Db 115 YNKPGEDEVVYAKQTL---EKVPLQ-KIESMPKELELEPYTAKGCKKORAPATPKSSS 169
 QY 119 GGCWQREDSGDAAHAFKSPSKENKK 146
 Db 170 GGA--GASTGSGTSSAAMVTSQSGSGSTK 195

RESULT 11
 GCNS_YEAST STANDARD; PRT: 439 AA.
 ID GCNS_YEAST
 AC 003330;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Transcriptional activator GCNS.
 GN GCNS OR ADA4 OR YGR252M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93011009; PubMed=1396595;
 RA Georgakopoulos T., Thireos G.;
 RT "Two distinct yeast transcriptional activators require the function
 RT of the GCNS protein to promote normal levels of transcription."
 RL EMO J. 11:415-4152(1992).
 RN [2]
 RP SEQUENCE OF 1-170 FROM N.A.
 RC STRAIN=8288C / FY1679;
 RX MEDLINE=9279234; PubMed=9133742;
 RA Petrolli F., Garignani G., Pavanello A., Guerreiro P., Azevedo D.,
 RA Rodriguez-Pousada C., Melchiorretto P., Panzeri L.,
 RA Agostoni Carbone M.L.;
 RT "Analysis of a 17.9 kb region from Saccharomyces cerevisiae
 RT chromosome VII reveals the presence of eight open reading frames,
 RT including BRF1 (TFR1B70) and GCNS genes."
 RL Yeast 13:373-377(1997).
 RN [3]
 RP SEQUENCE OF 170-439 FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=9279233; PubMed=9133741;
 RA Mazzoni C., Ruzzi M., Rinaldi T., Solinas F., Montedove F.,
 RA Frontali L.;
 RT "Sequence analysis of a 10.5 kb DNA fragment from the yeast
 RT chromosome VII reveals the presence of three new open reading frames
 RT and of a tRNAIle gene."
 RL Yeast 13:369-372(1997).
 RN [4]
 RP ASSOCIATION WITH ADA2.
 RX MEDLINE=95045371; PubMed=7957049;
 RA Marcus G.A., Silverman N., Berger S.L., Horluchi J., Guarente L.;
 RT "Functional similarity and physical association between GCNS and
 RT ADA2: putative transcriptional adaptors."
 RL EMO J. 13:4807-4815(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 99-262.
 RX MEDLINE=99362688; PubMed=10430873;
 RA Trivelp R.C., Rojas J.R., Steiner D.E., Venkataramani R.N., Wang L.,
 RA Zhou J., Allis C.D., Berger S.L., Marmorstein R.;
 RT "Crystal structure and mechanism of histone acetylation of the yeast
 RT GCNS transcriptional coactivator."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:8931-8936(1999).
 CC -1- FUNCTION: GENERAL TRANSCRIPTIONAL ACTIVATOR OPERATING IN CONCERT
 CC WITH CERTAIN OTHER DNA-BINDING TRANSCRIPTIONAL ACTIVATORS SUCH AS

CC GCN4 OR HAP2/3/4. FUNCTION AS AN HISTONE ACETYLTRANSFERASE (HAT)
 CC TO PROMOTE TRANSCRIPTIONAL ACTIVATION. HAS A STRONG PREFERENCE FOR
 CC LYSINE 14 OF H3 AND A SOMEWHAT LOWER PREFERENCE FOR LYSINES 8 AND
 CC 16 OF HISTONE H4.
 CC -1- SUBUNIT: HETERO-DIMER WITH ADA2. PART OF THE ADA/GCN5 COMPLEX THAT
 CC CONSISTS OF HELL/ADA1, ADA2, ADA3, SPT20/ADA5 AND GCN5.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68628; CAA48602.1; -
 DR EMBL: Z73037; CAA97281.1; -
 DR EMBL: X99228; CAA67614.1; -
 DR PIR: S28051; S28051.
 DR PDB: 1YGH; 02-AUG-99.
 DR TRANSFAC; T02145; -
 DR SGD: S0003484; GCN5.
 DR InterPro: IPR00182; Acetyltransf_GCN5.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00583; Acetyltransf_1.
 DR Pfam: PF00439; Bromodomain; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR Transcription regulation; DNA-binding; Activator; Trans-acting factor;
 KW 3D-structure.
 FT ACT_SITE 173 173 GENERAL BASE.
 FT DOMAIN 344 414 BROMODOMAIN.
 SQ SEQUENCE 439 AA; 51069 MW; 3200730DDC7EE70D CRC64;
 Query Match 15.2%; Score 130; DB 1; Length 439;
 Best Local Similarity 40.6%; Pred. No. 0.0013;
 Matches 28; Conservative 11; Mismatches 30; Indels 0; Gaps 0;
 QY 6 FSPFVDTAPGYSMITIKHPDMFSTMKERIKNNDYOSIELDKNFKMTNMTYKPEP 65
 DB 352 FLOPVKEKVPDYDFIKPEPMDLSTWEIKLESNKYQKMEFDYDAIVNNGCMYNGENT 411
 QY 66 IYYKAKKL 74
 DB 412 STYKIANRL 420
 RESULT 12
 CBP_MOUSE STANDARD; PRT; 2441 AA.
 AC P45481;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CREB-binding protein.
 GN CREBBP OR CBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94019866; PubMed=8413673;
 RA Christia J.C., Kwok R.P.S., Lamb N., Haglwaara M., Montminy M.R.,
 RA Goodman R.H.;
 RT "Phosphorylated CREB binds specifically to the nuclear protein CBP";
 RL Nature 365:855-859(1993).

CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSIVE GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S66385; AAB28651.1; -
 DR TRANSFAC; T01318; -
 DR MGI: MGI:1098280; Crebpb.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR003101; KIX.
 DR InterPro: IPR000197; TAZ_finger.
 DR InterPro: IPR000433; ZnF_Z2.
 DR Pfam: PF00439; Bromodomain; 1.
 DR Pfam: PF02172; KIX; 1.
 DR Pfam: PF02135; ZF-TAZ; 2.
 DR Pfam: PF00569; Z2; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00291; ZnF_Z2; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS01357; ZF_Z2_1; 1.
 DR PROSITE; PS50135; ZF_Z2_2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KW Zinc-finger.
 FT DOMAIN 1104 1176 BROMODOMAIN.
 FT ZN_FING 1702 1745 Z2-TYPE.
 FT DOMAIN 1062 1065 POLY-GLU.
 FT DOMAIN 1556 1563 POLY-GLU.
 FT DOMAIN 1944 1949 POLY-PRO.
 FT DOMAIN 1968 1971 POLY-GLN.
 FT DOMAIN 2082 2086 POLY-GLN.
 FT DOMAIN 2200 2216 POLY-GLN.
 FT DOMAIN 2296 2299 POLY-GLN.
 SQ SEQUENCE 2441 AA; 265474 MW; 0AB8028C3112F419 CRC64;
 Query Match 15.1%; Score 129; DB 1; Length 2441;
 Best Local Similarity 27.0%; Pred. No. 0.012; Mismatches 69; Indels 44; Gaps 9;
 Matches 51; Conservative 25;
 QY 1 DPSAF-FSPFVDTI--APGYSMITIKHPDMFSTMKERIKNNDYOSIELDKNFKMTNNA 57
 DB 1106 DPESLPFRQPVDPQLIGIDYDIYKNPMDLSTIKRKLDTGQYQEPWQYVVDVRLMFNNA 1165
 QY 58 MIYNKPEITYYKAAKLLSGMKILISQERISQSIDPMADLQTRKCKDGTDT----- 112
 DB 1166 WLYNRTKTSRYKFCSKL-----AEVEFOE-IDPMOSLIGCCG---RKYEFPQTLCCYG 1216
 QY 113 -----SOSGEDGCGW---QREDESGDAEAAFAKSPSKENKKKDKMDLE 153
 DB 1217 KOLCTIPRDAAYSYONRHFQCKFTETIQGENVILGD-----DPSQPTITTSKDPFE 1269
 QY 154 DKFSNNLE 162
 DB 1270 KK-KNDTLD 1277
 RESULT 13
 BRD4_HUMAN STANDARD; PRT; 1362 AA.
 ID BRD4_HUMAN
 AC 060885; Q96PD3;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Bromodomain-containing protein 4 (HUNK1 protein).
 GN BRD4 OR HUNK1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA French C.A., Fletcher J.A.;
 RT "Human BRD4 protein".
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE OF 1-722 FROM N.A.
 RC TISSUE=Placenta;
 RA Weber B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF386649; AAL26987.1; -
 DR EMBL: Y12059; CA72780.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 2.
 DR PROSITE: PS00633; BROMODOMAIN_1; 1.
 DR PROSITE: PS0014; BROMODOMAIN_2; 2.
 KM Bromodomain; Repeat; Nuclear protein.
 FT DOMAIN 75 147 BROMODOMAIN 1.
 FT 368 440 BROMODOMAIN 2.
 FT 535 594 LYS-RICH.
 FT 692 717 SER-RICH.
 FT 703 714 POLY-SER.
 FT 738 743 POLY-HIS.
 FT 757 761 POLY-PRO.
 FT 764 770 POLY-GLN.
 FT 771 775 POLY-PRO.
 FT 776 783 POLY-PRO.
 FT 954 964 POLY-PRO.
 FT 974 986 POLY-PRO.
 FT 1011 1014 POLY-PRO.
 FT 1028 1033 POLY-PRO.
 FT 1283 1300 POLY-GLN.
 FT 1301 1308 POLY-ALA.
 FT 1335 1338 POLY-ARG.
 FT 720 721 EM -> GP (IN REF. 2).
 SO SEQUENCE 1362 AA; 152219 MW; D52FEFICF960907 CMC64;

Query Match 14.8%; Score 126.5; DB 1; Length 1362;
 Best Local Similarity 21.2%; Pred. No. 0.0094;
 Matches 43; Conservative 33; Mismatches 52; Indels 75; Gaps 5;

QY 18 YSMITKHPDFTSMKEIKKNDYQSEIEELKDNFKLMCTNMTINKPETIYKAKAKU--- 74
 DB 390 YCDIIRKHPDMSTIKSKLEAREYRDAQEFADAVRLMFMSNCKYKPNPDHEVVAMARKLQDV 449
 QY 75 -----LHSGMKILTS----- 83
 DB 450 FEMRPAKMPDEPPEPVAVSSPAVPPPTKVVAVAPSSSDSSSDSSSDSSDSTDSSEERRAQ 509
 QY 84 -----QERISLQKSIDFMADLQKTRKQKDGDTDSGEGDGCQWQREEDSGDAEHAFAK 138

DB 510 RLAELEQKVAHVHQALASQPQONKPKKEKD-----KKEKK-----KEKHRRK 554
 QY 139 SPSEKNNK-KDKMDLEDKFSKN 160
 DB 555 EEVEENKSKAKPEPPKTKKNN 577
 RESULT 14
 ID BRM_DROME STANDARD; PRT; 1638 AA.
 AC P25439; O9YUW5; O9YUW6;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Homeotic gene regulator (Brahma protein).
 GN BRM OR CG5942 OR CG18438.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE=92154670; PubMed=1346755;
 RA Tamkun J.W., Deuring R., Scott M.P., Kissinger M., Pattatucci A.M.,
 RA Kautman T.C., Kennison J.A.;
 RT "Brahma: a regulator of Drosophila homeotic genes structurally
 RT related to the yeast transcriptional activator SNF2/SWI2.";
 RL Cell 68:561-572(1992).
 RN (2)
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazer J.G., Champe M., Pfeiffer B.D.,
 RA Abrial J.F., Adysayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: MAY ACT AS COACTIVATOR, ASSISTING ONE OR MORE DEDICATED

CC TRANSCRIPTIONAL ACTIVATORS OF ANT-C AND BX-C HOMEOTIC GENES. CAN
 CC CONTACT THE REPRESSIVE EFFECT OF POLYCOMB PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN UNFERTILIZED EGGS AND
 CC EARLY EMBRYOS.
 CC -1- MISCELLANEOUS: 'BRAHMA' MEANS 'FATE' IN INDIA.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SMF2/RAD54 HELICASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M85049; AAA19661.1; -
 CC EMBL; AE003529; AAF49557.1; -
 CC EMBL; AE003529; AAF49558.2; -
 CC PIR; A42091; A42091.
 CC
 CC Flybase; FBgn0000212; btm.
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR000330; SMF2_N.
 CC Pfam; PF00439; bromodomain; 1.
 CC Pfam; PF00271; helicase_C; 1.
 CC Pfam; PF00176; SMF2_N; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00287; BROMO; 1.
 CC SMART; SM00490; HELIC; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; 1.
 CC PROSITE; PS00633; BROMODOMAIN_2; 1.
 CC KMW Actinuator; Bromodomain; Transcription regulation; Nuclear protein;
 CC KMW Actinuator; Bromodomain; Helicase; ATP-binding; Alternative splicing.
 CC FT NP_BIND 201 390 GIN/PRO-RICH.
 CC FT SITE 798 805 ATP (POTENTIAL).
 CC FT DOMAIN 1385 1392 DEGH BOX.
 CC FT DOMAIN 1394 1404 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 1405 1410 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 1415 1432 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 1443 1513 ARG/LYS-RICH (BASIC).
 CC FT DOMAIN 1631 1638 BROMODOMAIN.
 CC FT VARSPIC 121 121 G->A (IN SHORT ISOFORM).
 CC FT VARSPIC 122 222 MISSING (IN SHORT ISOFORM).
 CC FT CONFLICT 687 687 D->Y (IN REF. 1).
 CC SQ SEQUENCE 1638 AA; 18508 MW; A4494B29FAF2EA2A CRC64;

Query Match 14.5%; Score 124; DB 1; Length 1638;
 Best Local Similarity 28.8%; Pred. No. 0.019;
 Matches 40; Conservative 21; Mismatches 60; Indels 18; Gaps 4;

ID CBP_HUMAN STANDARD; PRT; 2442 AA.
 AC 092793; Q16376; Q00147;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CREB-binding protein.
 GN CREBBP OR CBP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_Taxid=9606;
 RX MEDLINE=97385172; PubMed=9238046;
 RA Schlobo O.M., Borrow J., Tomek R., Reshmi S., Harden A.,
 RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
 RA Zelenik-Le N.J.;
 RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
 RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97321049; PubMed=9177780;
 RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
 RA Lushnikova T., van Omme G.J.B., Goodman R.H., Deaven L.L.,
 RA Doggett N.A., Peters D.J.M., Breuning M.H.;
 RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
 RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
 RT 16p13.3.";
 RL Genomics 42:96-144(1997).
 RN [3]
 RP SEQUENCE OF 1-405 FROM N.A.
 RX MEDLINE=96376968; PubMed=8782817;
 RA Borrow J., Stanton V.P., Andersen J.M., Becher R., Behn F.G.,
 RA Chaganti R.S.K., Civan C.I., Distcheche C., Dube I., Fritschauf A.M.,
 RA Housman D., Mittleman F., Volinia S., Watson D.E.;
 RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
 RT a putative acetyltransferase to the CREB-binding protein.";
 RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP ARGUMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSIVE GENES.
 CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(q23;p13.3)
 CC INVOLVING CBP AND MLL.
 CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
 CC SYNDROME (RTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
 CC ABNORMALITIES, BROAD THUMBES, BROAD BIG TOES, MENTAL RETARDATION
 CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U47741; AAC51770.1; -
 CC EMBL; U85962; AAC51331.1; -
 CC EMBL; U89354; AAC51339.1; -
 CC EMBL; U89355; AAC51340.1; -
 CC MIM; 600140; -
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR003101; KIX.
 CC InterPro; IPR000197; TAZ_finger.
 CC InterPro; IPR000433; ZnF_ZZ.
 CC Pfam; PF00439; bromodomain; 1.

RESULT 15
 CBP_HUMAN

DR Pfam; PF02172; KIX; 1.
DR Pfam; PF02135; zf-TAZ; 2.
DR Pfam; PF00569; ZF; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00291; Znf_Z2; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01357; ZF_Z2_1; 1.
DR PROSITE; PS50135; ZF_Z2_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Chromosomal translocation; Zinc-finger.
FT ZN_FING 1701 1744
FT DOMAIN 363 430
FT DOMAIN 452 683
FT DOMAIN 1103 1175
FT DOMAIN 1061 1064
FT DOMAIN 1199 1487
FT DOMAIN 1555 1562
FT DOMAIN 1675 1849
FT DOMAIN 1943 1948
FT DOMAIN 1967 1970
FT DOMAIN 2081 2085
FT DOMAIN 2199 2216
FT DOMAIN 2245 2248
FT DOMAIN 2297 2300
FT CONFLICT 1511 1513
FT CONFLICT 1724 1725
FT CONFLICT 1770 1770
FT CONFLICT 1789 1789
FT CONFLICT 1812 1812
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;
ED -> VV (IN REF. 2).
V -> L (IN REF. 2).
N -> F (IN REF. 2).
T -> P (IN REF. 2).

Query Match 14.4%; Score 123; DB 1; Length 2442;
Best Local Similarity 34.3%; Pred. No. 0.036;
Matches 34; Conservative 16; Mismatches 41; Indels 8; Gaps 4;

OY 1 DPSAF-FSEPTDFI--APGYSMTIKHPMDSTMEKIKNNNDYQSIIEIKDNFKLMCTNA 57
Db 1105 DPESLPFRQPDQPLGIPDFEDIVKNPMDISTIKRKLDYGOYQYQYDDVWLMFNNA 1164
OY 58 MIYKRPETIYKAAKKLLHSGMKILISOERIOSLKSIDF 96
Db 1165 WLYNKRKTSRYVKFCSKL---AEVFQEI-IDPVNQSLGY 1198

Search completed: July 11, 2002, 15:55:26
Job time: 349 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:56:16 ; Search time 49.31 Seconds

(without alignments)
571.855 Million cell updates/sec

Title: US-09-687-230-2_COPY_151_313

Perfect score: 855
Sequence: 1 DPSAFSEFPYDTFAPGYSM.....NKKKDKMDLEKSKNNLER 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_YEAST:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	855	100.0	651	4 Q9NP11	Q9NP11 homo sapien
2	855	100.0	652	4 Q9UH59	Q9UH59 homo sapien
3	777	90.9	651	11 O88665	O88665 mus musculu
4	610	71.3	459	4 Q9BV48	Q9BV48 homo sapien
5	375	43.9	501	4 Q9H8M2	Q9H8M2 homo sapien
6	328	38.4	233	4 Q9H7R9	Q9H7R9 homo sapien
7	284.5	33.3	861	5 Q9VIX2	Q9VIX2 drosophila
8	200.5	23.5	636	5 Q17581	Q17581 caenorhabd1
9	174.5	20.4	952	10 Q9SNT9	Q9SNT9 oryza sativ
10	174	20.4	715	4 Q9V4O3	Q9V4O3 homo sapien
11	174	20.4	805	4 Q95692	Q95692 homo sapien
12	172.5	20.2	706	11 Q9SVV4	Q9SVV4 mus musculu
13	169.5	19.8	652	10 Q9LMB8	Q9LMB8 arabidopsis
14	166	19.4	145	10 Q9FIA2	Q9FIA2 arabidopsis
15	163.5	19.1	556	10 Q9SEF1	Q9SEF1 arabidopsis
16	160	18.7	1430	5 Q9V4J4	Q9V4J4 drosophila

17	156	18.2	473	5 Q9Y0F2	Q9Y0F2 toxoplasma
18	156	18.2	1169	5 Q9NTS0	Q9NTS0 toxoplasma
19	155.5	18.2	1568	13 Q90755	Q90755 gallus gall
20	151.5	17.7	757	4 Q15355	Q15355 homo sapien
21	149.5	17.5	418	5 Q27198	Q27198 tetrahymena
22	148.5	17.4	236	11 Q9D007	Q9D007 mus musculu
23	148.5	17.4	454	3 Q9HUK2	Q9HUK2 schizosach
24	148.5	17.4	1235	4 Q9H0E9	Q9H0E9 homo sapien
25	147	17.2	513	5 Q9VCG6	Q9VCG6 drosophila
26	147	17.2	1630	13 Q90753	Q90753 gallus gall
27	146.5	17.1	729	13 Q90971	Q90971 gallus gall
28	146.5	17.1	733	13 Q73897	Q73897 gallus gall
29	146	17.1	920	4 Q43178	Q43178 homo sapien
30	146	17.1	920	4 Q969M9	Q969M9 homo sapien
31	146	17.1	2065	5 Q97068	Q97068 drosophila
32	144	16.8	1022	11 Q63928	Q63928 mus sp. brg
33	144	16.8	1647	4 Q9HBD3	Q9HBD3 homo sapien
34	144	16.8	1654	5 Q9VC36	Q9VC36 drosophila
35	144	16.8	1679	4 Q9HBD4	Q9HBD4 homo sapien
36	142	16.6	726	11 Q9J125	Q9J125 mus musculu
37	142	16.6	754	4 Q15310	Q15310 homo sapien
38	142	16.6	801	4 Q969U4	Q969U4 homo sapien
39	141.5	16.5	3080	5 Q9YRY3	Q9YRY3 drosophila
40	140	16.4	374	5 Q76561	Q76561 caenorhabd1
41	139	16.3	1792	5 Q9YU19	Q9YU19 caenorhabd1
42	138.5	16.2	798	11 Q54795	Q54795 mus musculu
43	138.5	16.2	798	11 Q88411	Q88411 mus musculu
44	135.5	15.8	1865	11 Q60544	Q60544 mesocricetu
45	135	15.8	1937	5 Q9W3L3	Q9W3L3 drosophila

ALIGNMENTS

RESULT 1
ID Q9NP11 PRELIMINARY: PRT: 651 AA.
AC Q9NP11;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE BROMODOMAIN CONTAINING PROTEIN (NAG4).
GN BP465.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=FETAL BRAIN.
RC TISSUE=FETAL BRAIN.
RA Kzyshkowska J.G., Dobner T.G.;
RT "Cloning of a human bromodomain containing protein.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA YU Y., LI G.Y.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0271881; CAB72445.1; -
DR EMBL; AF152604; AAF75126.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PSS0014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74138 MW; 29B7947644C215E7 CRC64;

Query Match 100.0%; Score 855; DB 4; Length 651;
Best Local Similarity 100.0%; Pred. No. 2.8e-63;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSAFSEFPYDTFAPGYSMLIKHPMDFSTMKERIKNNYOSIELDKNFKMTNMIY 60
|||||

```

Db 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 211 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 270
OY 121 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
DB 271 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 2
O9UH59 PRELIMINARY; PRT; 652 AA.
AC 09UH59;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE BROMODOMAIN PROTEIN CELTIX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Stahl A., Enserink J., Stein J.L., Stein G.S., van Wijnen A.J.;
RT "Molecular analysis of the human bromodomain protein Celtix-1."
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF213969; AAF19526.1;
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 652 AA; 74084 MW; 70F2B654B2618529 CRC64;

Query Match 100.0%; Score 855; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.8e-63;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 60
DB 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 211 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 270
OY 121 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
DB 271 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 3
O88665 PRELIMINARY; PRT; 651 AA.
AC 088665;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE BROMODOMAIN-CONTAINING PROTEIN BP75.
CN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BAIB/C; TISSUE=BRIN;
RC MEDLINE=99456832; PubMed=10526152;
RA Cuppen E., van Ham M., Peeters B., Wieringa B., Hendriks W.;
RT "Identification and molecular characterization of BP75, a novel

```

```

RT bromodomain-containing protein."
RL FEBS Lett. 459:291-298(1999).
DR EMBL; AF084259; AAC33302.1;
DR MGI; MGI:1349766; Brd7.
DR InterPro: IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 651 AA; 74000 MW; 5D34B4F14FD51350 CRC64;

Query Match 90.9%; Score 777; DB 11; Length 651;
Best Local Similarity 89.6%; Pred. No. 8.7e-57;
Matches 146; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

OY 1 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 60
DB 151 DSAFFSPVTDPIAGYSMIKHHPDFTMKKIKNNNDYOSIEELKDNFKLMCTNAMY 210
OY 61 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 120
DB 211 NKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRKOKDGTDSQSGEDG 270
OY 121 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
DB 271 CWOEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 313

RESULT 4
O9BV48 PRELIMINARY; PRT; 459 AA.
AC 09BV48;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE SIMILAR TO BROMODOMAIN-CONTAINING 7 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=BRIN; NEUROBLASTOMA;
RC Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC001611; AAH01611.1;
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain; 1.
FT NON-TER
SQ SEQUENCE 459 AA; 51714 MW; 9EF21D1454BCE63B CRC64;

Query Match 71.3%; Score 610; DB 4; Length 459;
Best Local Similarity 98.3%; Pred. No. 4.8e-43;
Matches 116; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 46 LKDNFKLMCTNAMYNKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRK 105
DB 3 IQDNFKLMCTNAMYNKPEITYYKAAKLLHSGMKILISOERIOSLQSIDFMADLOKTRK 62
OY 106 OKDGTDSQSGEDGCGWOREEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 163
DB 63 OKDGTDSQSGEDGCGWOREEREDSGDAEAHAFKSPSKENKKDKMDLEDFKSNLNR 120

RESULT 5
O9H8M2 PRELIMINARY; PRT; 501 AA.
AC 09H8M2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE CDNA FLJ13441 FIS, CLONE PLACE1002775, WEAKLY SIMILAR TO

```

DE PEREGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RI EMBL: AK023503; BABI4591.1;
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO.1.
 DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 501 AA; 55664 MW; 5C68F53097BA073C CRC64;

Query Match 43.98; Score 375; DB 4; Length 501;
 Best Local Similarity 47.58; Pred. No. 2e-23;
 Matches 77; Conservative 28; Mismatches 37; Indels 20; Gaps 3;

QY 1 DPSAFSPVPTDFIAPGYSMIIKHPDFTMKERIKNNQSYIELDNFKLCTNMATY 60
 DB 40 DPHGFAPVPTDAIAPGSMIIKHPDFTMKERIKNNQSYIELDNFKLCTNMATY 99
 QY 61 NKPEITYYKAAKLLHSGMKILSOERISLQSIDFMADIKTRKOKDGTDSQSGEDG 120
 DB 100 NRPTVYVYKLAKKILHAGFKMSKERLLALKRSMSPQDM-----DFSQAALLG 149
 QY 121 CQORERENSGAEAAHAFSP-----SKNRKKDKMDLEDKR 157
 DB 150 -----NEDTAVEEPYEPVPOVETAKSKRPSREVISCMEF 186
 RESULT 6
 Q9H7R9 PRELIMINARY; PRT; 233 AA.
 AC Q9H7R9;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CDNA FLJ14330 FIS, CLONE PLACE4000261, WEAKLY SIMILAR TO
 DE PEREGRIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RI EMBL: AK024392; BABI4907.1;
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO.1.
 DR PROSITE: PSS0014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 233 AA; 26233 MW; FB96245BD86F4EB CRC64;

Query Match 38.48; Score 328; DB 4; Length 233;
 Best Local Similarity 67.18; Pred. No. 7.1e-20;
 Matches 57; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

QY 1 DPSAFSPVPTDFIAPGYSMIIKHPDFTMKERIKNNQSYIELDNFKLCTNMATY 60
 DB 40 DPHGFAPVPTDAIAPGSMIIKHPDFTMKERIKNNQSYIELDNFKLCTNMATY 99
 QY 61 NKPEITYYKAAKLLHSGMKILSOE 85
 DB 100 NRPTVYVYKLAKKILHAGFKMSKQ 124
 RESULT 7
 Q9VLX2 PRELIMINARY; PRT; 861 AA.
 AC Q9VLX2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CG7154 PROTEIN.
 GN CG7154.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Artl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Borokova D., Botchan M.R., Bouck B.P., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Daventport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jengam C.,
 RA Jaitani M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.C., Scheeler F., Smith H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RI EMBL: AE003618; AAF52557.1;
 DR Flybase: FBgn0031947; CG7154.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain.1.

DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 861 AA; 95922 MW; E27D528E5F9B3A3F CRC64;

Query Match 33.3%; Score 284.5; DB 5; Length 861;
 Best Local Similarity 41.0%; Pred. No. 1.2e-15;
 Matches 68; Conservative 30; Mismatches 53; Indels 15; Gaps 4;

OY 1 DPSAFSPVDTFIAGYSMIKHPMDFSTMEKIKNNNDYOSIEELKDNFKLMTCTNAMY 60
 DB 295 DPHQFFAMPTDMDAPGYSSIIISRPDMFSTMRKIDHETALTEFTDCKLCEANAIKY 354
 OY 61 NKPEITYYAAKAKLHSGMKILSOERI-OSLKOSIDPMADLQKTRKOKDGTDSGSDG 119
 DB 355 NHVDYTYNNAKAKLQVGMKHLQPENIMRSILKPLSGYMERL---TARELGFELSSND--- 408
 OY 120 GCMQREDESDG-----AEAHAFKSPSKENKKDKMDLQKTRKOKDGTDSGSDG 159
 DB 409 --MSRENNDSADBGASTGAEEPTPTPAOLEEERKRTLRLENAKPTH 452

RESULT 8
 ID 017581 PRELIMINARY; PRT; 636 AA.
 AC 017581;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE C0146.7 PROTEIN.
 GN C0146.7
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
 OC Rhabdilitida; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Berts M.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99069613; PubMed-9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; 271258; CAA95779.1;
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 1.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 636 AA; 71339 MW; 02A5B089B8DCD663 CRC64;

Query Match 23.5%; Score 200.5; DB 5; Length 636;
 Best Local Similarity 31.6%; Pred. No. 8.9e-09;
 Matches 54; Conservative 32; Mismatches 36; Indels 29; Gaps 5;

OY 1 DPSAFSPVDTFIAGYSMIKHPMDFSTMEKIKNNNDYOSIEELKDNFKLMTCTNAMY 60
 DB 167 DPEQYFAFPVPSMAPDYRDIKTPMDQITRENEDGKYASLPAMKEDCELIVSNAYOY 226
 OY 61 NKPEITYYAAKAKLHSGMKILSOERIOSLKOSIDF-----MADLQKTR--- 104
 DB 227 NOPNTVFYLAARLSNLAAYFGEOYLRFSLPMAKIPFEIYGIPLAVPERMTN 286
 OY 105 KOKDGTDSGSDGCGWQ-----REREDSGDAEAHAFKSPSKENKKKK 149
 DB 287 KRKAAYKRGMTSED--CLQVADPKYERLSAKLPEAN-----NPKNKKMK 330

RESULT 9

O95N19
 ID 095N19 PRELIMINARY; PRT; 952 AA.
 AC 095N19;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SIMILAR TO DJS22J7.2.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 3, PAC
 RT clone: P0043E01."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP000615; BAA85417.1;
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 952 AA; 105312 MW; 91D8E06A5A1769D CRC64;

Query Match 20.4%; Score 174.5; DB 10; Length 952;
 Best Local Similarity 36.8%; Pred. No. 2e-06;
 Matches 42; Conservative 21; Mismatches 36; Indels 15; Gaps 3;

OY 1 DPSAFSPVDTFIAGYSMIKHPMDFSTMEKIKNNNDYOSIEELKDNFKLMTCTNAMY 60
 DB 340 DTGYGSEVVDHEHLVDYHELIEHMFSTIEKLNLSYTLLEQFENDVFLTSNASTY 399
 OY 61 NKPEITYYAAKAKLHSGMKILSOERIOSLKOSIDPMADLQKTRKOKDGTDSQ 114
 DB 400 NSDDTVYYRQL-----VGWK---TERIVSVRSQ-----QKRIPIPGSDMSR 438

RESULT 10
 ID 09Y403 PRELIMINARY; PRT; 715 AA.
 AC 09Y403;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOHETICAL 79.7 KDA PROTEIN (FRAGMENT).
 GN DKE2P434B094.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Wambutt R., Heupner D., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL080149; CAA45742.1;
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00855; PWM; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00293; PWM; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 715 AA; 79702 MW; D397FF5501ED9990 CRC64;

```

Query Match      20.4%; Score 174; DB 4; Length 715;
Best Local Similarity 38.4%; Pred. No. 1.6e-06;
Matches 33; Conservative 22; Mismatches 31; Indels 0; Gaps 0

OY 1 DPSAFSEPVTDIFAPGYSMIIRKHPMDSTSMKEIKINDYOSIEELDKNFILMCNTNAMY 60
   || : | : | : | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 108 DPAAHFAQVSLAEVPDYLDHIKHPMDFATMRKRRLAEGTKNLHEFEDEDLIIDCMCKI 167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 61 NKPEITYYKAARKKLHSGMKLISQER 86
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 168 NARDTVFYRAAVLRIDGGVLEROAR 193
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 11
095692 PRELIMINARY; PRT; 805 AA.
AC 095692;
DT 01-MAY-1999 (TREMBLrel, 10, Created)
DT 01-MAY-1999 (TREMBLrel, 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE DJS24E15.1 (PEREGRIN (BR140 PROTEIN)) (FRAGMENT).
GN DJS24E15.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Palmer S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; 284485; CAB06488.1; -
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR000313; PWMP.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PR00628; PHD; 1.
DR Pfam; PR00855; PWMP; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00293; PWMP; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
DR NON_TER 1
FT FT 805
SQ SEQUENCE 805 AA; 90851 MW; E28C017F5C545334 CRC64;

Query Match      20.4%; Score 174; DB 4; Length 805;
Best Local Similarity 31.2%; Pred. No. 1.9e-06;
Matches 44; Conservative 27; Mismatches 56; Indels 14; Gaps 3;

OY 1 DPSAFSEPVTDIFAPGYSMIIRKHPMDSTSMKEIKINDYOSIEELDKNFILMCNTNAMY 60
   || : | : | : | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 524 DPAAHFAEPVN---LNYLEFLSKPMDSTMRKRKLESHTLYTLEFEEDFMIIYTNCMKY 579
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 61 NKPEITYYKAARKKLHSGMKLISQERISLKOSID----FMADLOKTROKDGSTDTSQS 115
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 560 NAKDTIFRAAVRLLEDGAILRRARRQAENIGYDPERGTHLPDGOHNPREDPGFPVRG 639
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 116 GEDGCGMQRERESGDAAEAHA 136
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 640 AAEAGAA-----GETGPGERHA 655
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
099JV4 PRELIMINARY; PRT; 706 AA.
AC 099JV4;
DT 01-JUN-2001 (TREMBLrel, 17, Created)
DT 01-JUN-2001 (TREMBLrel, 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE HYPOHETICAL 79.9 KDA PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
```

OC Eukaryot; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC005647; AAH05647.1; -
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR000313; PWM.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00855; PWM; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00293; PWM; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SO SEQUENCE 706 AA; 79852 MW; 87C549BB591B0639 CRC64;

```

Query Match          20.2%; Score 172.5; DB 11; Length 706;
Best Local Similarity 34.7%; Pred. No. 2.1e-06;
Matches 41; Conservative 20; Mismatches 56; Indels 1; Gaps 1

OY 1 DPSAFSEFPVTDIFAPGYSMIIRKHPDFTSMKEKIKNNDYOSIELDKNFKMCTNAMY 60
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 141 DTGNISEFVPLSEVPDYLDHIKKPMDFFIMKKOLEAVRYLRNFDPEEDENLIYSNCKLY 200
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 61 NKPEETIYKAAKKILHSGMKILISOERQSLKOSIDFPADIDQKTRKQKQDITDSOSGED 118
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 201 NAKDTTFYKAAVRLREOGAVLQARQARQAKMGIDFETGMIPHNLA-GDEVSHHTED 257
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 13
OQ1LM88
ID OQ1LM88 PRELIMINARY; PRT; 652 AA.
AC O9LM88:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
F2D10.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RN
RP SEQUENCE FROM N.A.
RA Shim P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bel Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayerl A.,
RA Tortumt J.R., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
Ecker J.R.;
"Genomic sequence for Arabidopsis thaliana BAC F2D10 from chromosome
1."
RT
RT
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL, AC069251, AAF80635.1; -.
DR InterPro, IPR001487; Bromodomain.
DR Pfam, PF00439; bromodomain.1.
DR PRINTS, PR00503; BROMODOMAIN.
DR SMART, SM00297; BROMO.1.
DR PROSITE, PS0014; BROMODOMAIN_2.1.
SQ
SEQUENCE 652 AA; 72956 MW; 35782E0E061D74C5 CRC64;

```

Db 191 DTYGYSDPVPPEELPDYFEIKNPMDSITRNKLDSCAYSTLEOFEDVFLICTNAMEY 250
 Oy 61 NKPEITYYKAKKILHSGMKILISOERIQSLKOSID 95
 Db 251 NSADTVYRQAR-----AIQELAKKDFENLRQDSD 280

RESULT 14

O9FIA2 PRELIMINARY; PRT; 145 AA.

AC O9FIA2? 16
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 17, Last sequence update)
 DE DBJ|BA05417.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RC MEDLINE=99397451; PubMed=10470850;
 RA Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
 RA MiyaJima N., Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
 RT Sequence features of the regions of 1,011,550 bp covered by seventeen
 RT P1 and TAC clones."
 RL DNA Res. 6:183-195(1999).
 DR EMBL: AB017059; BAB10578.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 145 AA; 16649 MW; BD3E5CEPF910CC63 CRC64;

Query Match 19.4%; Score 166; DB 10; Length 145;
 Best Local Similarity 45.3%; Pred. No. 1.3e-06;
 Matches 29; Conservative 16; Mismatches 19; Indels 0; Gaps 0;
 Oy 6 FSPVDTFLAPGYSMTIKRHPDSTKKEIKKNDYOSIEELKDNFKLMCTNMIYKPEP 65
 Db 67 YAEVPDEELPDYHDMIEHPDSTYRKILANGSYTLEELSDVLLCSNMQYNSDPT 126
 Oy 66 IYK 69
 Db 127 VYK 130

RESULT 15

O9SFX1 PRELIMINARY; PRT; 556 AA.

AC O9SFX1? 13
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOHETICAL 62.1 KDA PROTEIN.
 GN FLSM4.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Uteback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome 1 BAC FLSM4 genomic sequence."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC012394; AAF16663.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR Hypothetical protein.
 KW
 SQ SEQUENCE 556 AA; 62060 MW; 03878B1E71C891C2 CRC64;

Query Match 19.1%; Score 163.5; DB 10; Length 556;
 Best Local Similarity 27.6%; Pred. No. 9.3e-06;
 Matches 45; Conservative 28; Mismatches 55; Indels 35; Gaps 4;
 Oy 1 DPSAFSPVDTFLAPGYSMTIKRHPDSTKKEIKKNDYOSIEELKDNFKLMCTNMIYKPEP 56
 Db 161 DTYGYSDPADPEELPDYFEIKNPMDSITRNKLDSCAYSTLEOFEDVFLICTNAMEY 219
 Oy 57 AMIYKPEITYYKAKKILHSGMKILISOERIQSLKOSIDFMADLQTKRKORDGTDTDSQSG 116
 Db 220 AMEYNSADTVYRQARAMELELAKK-----DRGNLRQESDGEPEVSLIS 261
 Oy 117 EDGCGWQREDESDGDAEHAHAKSPKSKKKDKDMLERKFSN 159
 Db 262 QOPKVKYKRGK-----PPSGGLKQLEQSLIDRTTSD 292

Search completed: July 11, 2002, 15:56:17
 Job time: 360 sec


```

98 eutrhspTyrHisLeuLysLysLysIleLeuAlaArgLysAlaAla 114
   ::::: |||:::
435 CGAGATGAGACCGGGTGAGAAATGAGCAGAAAAAGATCTCCAGTGTCA 484
   ||||| ::::: |||::: |||::: |||:::
115 ArgAspIleGluLys...GluLysGluValGluProGluValGluGluGlu 130
485 CGGCCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 528
   ||||| ::::: |||::: |||::: |||:::
130 uValProLysGluProThrProProProProProProProProProPheS 147
   ||||| ::::: |||::: |||::: |||:::
529 GCTCTTATGACCAAAAGAGAGATAGACAGACACCCCTTCAAGAGACT 578
   ||||| ::::: |||::: |||::: |||:::
147 etser.....TyrLeuProIleGluLeuMet 155
579 TTGAATCACTGATGAGACAAATTCAGAGAAAAGATCAAGTCTTCTT 628
   ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
156 GluAspIleIleLeuArgLysLeuValGluLysAspProGluGluGlu 172
629 TTGATTTCCGTGACTGATTTTATGCTCTGCTGCTGCTGCTGCTGCTGCT 678
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 eAlaIleProValThrProSerMetAlaProAspTyrArgAspIleIleL 189
679 AACCCCAATGATTTTATGATGATGATGATGATGATGATGATGATGATGAT 728
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 ysthrProMetAspLeuGluThrIleArgGluAsnIleGluAspGlyLys 205
729 TATCACTCCATGAGAACTAAAGATTAACCTTCAACTAATGCTGCTACTAA 778
   ||| ||||| ::::: |||::: |||::: |||::: |||::: |||::: |||:::
206 TyrAlaSerLeuProAlaMetLysGluAspCysGluLeuIleValSerAs 222
779 TGCCATGATTTACATAAACCCAGACGATTTTATTTAAAGCTGCAGAGA 828
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 nAlaIleGluIleTyrAsnGlnProAsnThrValPheTyrLeuAlaIleLys 239
829 AGCTGTGACTCAGGAATGAAATTTCTTACGAGAGAAAGATTCAGAGC 878
   ::||| ::::: |||::: |||::: |||::: |||::: |||::: |||:::
239 rgluSerAsnLeuIleAlaTyrTyrPheGlyGluGluIleTyrLeuAlaPhe 255
879 CTGAGACAGAGCATGACATTC..... 899
   ||| ||||| ::::: |||::: |||::: |||::: |||::: |||::: |||:::
256 LeuPheIleSerLeuProMetAlaAsnLysIleProPheGluIleValGlu 272
900 .....ATGGCTGACTGCAGAAAACCTGA.....AACCA 930
   ::||| ::::: |||::: |||::: |||::: |||::: |||::: |||:::
272 ylleArgProLeuAlaProValProLysGluArgThrMetAsnLysArgL 289
931 AAGATGGAACAGACACCTCAGAGAGTGGGAGAGCAGAGCTGCTGGCAG 980
   || ::::: |||::: |||::: |||::: |||::: |||::: |||:::
289 yAlaIleValLysAspGlyMetThrSerGluAsp.....CysLeuGln 303
981 .....AGAGAGAGAGGACTCTGGAGATGCCGAGAC 1012
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 ValAlaAspProLysValArgLysGluArgLeuSerAlaLysLeuProGluAl 320
1013 ACAGCCCTTCAAGAGTCCAGCAAGAAAGAAATTAAGAGAGCAAAAGATA 1062
   | ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
320 a.....AsnAsnProLysAsnLysLysM 328
1063 TCCCTGAGAGATGATTTAAACCAATTAATTTAGAGAGAGAGAGAGAGAG 1112
   || ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
328 etGlyLysLeuGlyPheLeuSer..... 335
1113 CTTGACCCGATGCTGAGAGATCTGAGAGAAAGCTGACAGCGGCTTGT 1162
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
336 .....GluLysAspGlyThrValIleuAsnValIleVal 346
1163 GAACAGTCAGTCCGAA.....TTGAAAGAGAGAGAGAGAGAGAGAGAG 1203
   | ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
346 lAlaIleLysAspSerGluAspGlyLysLeuGluAsnAlaProAlaArgArgV 363
1204 CAACGACGTTGGGACTTCTCCATCTCTGTGATGCCATTTGTAGAGAGAGCA 1253
   ::||| ::::: |||::: |||::: |||::: |||::: |||::: |||:::

```

```

363 alThrIleGlyAspIleValGlyProLeuGluGlu.....GlyThrPro 377
1254 GCGTACTGCGCTGCGAGACCTGGAGATGACAACTGGAAGCTGCTGG 1303
   ||| ::::: |||::: ||||| |||
378 Gly.....MetIleGlnMetAlaAspHisArgLeuPheSer.. 389
1304 AGTGAAATGCTTTGCGAGGGGTTCAAAAGAGATTAAGAGACAAAGACTCT 1353
   390 .....GlnAlaP 392
1354 CAGGTATATTTGAATTTATGGGCGCTACAGCTCTTATGACCGCATAT 1403
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 roValAlaAsnTyrLeuAsnTyrGlyProTyrSerSerPheAlaProMetTyr 408
1404 GACTCCACATTTGCAATATACAGAGAGATGATTTCTGATTTAATCTATTTC 1453
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 AspSerThrThrPalaThrMetThrLysGluAspThrAspLeuPheLeuArg 425
1454 AACCTATGGGGAAGACTGTGATCTTCCAAAGTGAATTCAGCATCCATGAGT 1503
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 gThrTyrGlyAspLysSerAsnAlaSerAspValMetSerMetArgArgP 442
1504 TTTTGGCCAGCTGCCAAGATTAATCCGATGTCATGCGACATGATTTACTG 1553
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 heValGlyAspCysProGluPheSerGluIleIleGly...SerLeuLeu 457
1554 GATGTTTAAACAAAGAGAGGAGCATTCAGAGACCTTCAAGAGATGAGAT 1603
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
458 AspThrLeuThrAspGlyGluHisSerLysThrMetLysGluLeuGlu.. 473
1604 GTCATTTGCTTGAAGATGAAAGGCCATCTAGAGACATTCACACAGAAAG 1653
   474 .....AsnAlaGlyLysG 478
1654 AATGGAGAGATTTACAGAGATGAGACCAACAGCGGCTTGGACCTCAGT 1703
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 luValLysGlu.....GluValAspAsnAspGluTyrLysAsnGluThr 492
1704 ACTCAAGACAGGCTCATAGCGCTGAAGAGAGATTAACAAATTTGGCGCTCC 1753
   ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
493 ValLeuSerLeuIleAspAspValSerSerIleSerAsnLeuGlyIleGlu 509
1754 AGTTGAAGTTTGTGACTCTGAAAGACTGAATATTTC..... 1790
   ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
509 uThrGlyPheLeuAsnAspIleArgGlnGlnValLeuValProAlaValGlu 526
1790 ..... 1790
526 luserAsnIleGluAsnAlaIleProGluPheMetAsnGluValAsnHis 542
1791 .....CAGAGAACTGATGAGACCAACAGATTCCTCAGGAGACT 1831
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
543 MetAsnValGlnGlnIleLeuAsnHisSerGlyGlnLysValLysAspRe 559
1832 CCAGAGAGCCCAAGATGAACTTTGAGCAGCAGAGCCCTTGGGAACATGA 1881
   | ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
559 uAlaHisIleGlnGlnLysArgLeuValGlnGlnProProMetIleM 576
1882 TCTGTCTTGGGCTCCATCAGA..... 1906
   ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
576 etserValGlnGlnLysIleGlnGlnLysLeuAlaIleLysLeu 592
1907 GAAATGACATCTTGTGAAACATGACCAATTAAT 1939
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
593 GlnGlnHisLeuAlaHisGlnMetThrThrHis 603
seq_name: p12534
seq_documentation_block:
hypotheical protein DKFZp434B094.1 - human (fragment)
C:Species: Homo sapiens (hmn)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000
C:Accession: T12534

```

R.Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: 217524
 A:Accession: T12534
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-715 <MW>
 A:Cross-references: EMBL:AL080149
 A:Experimental source: adult testis; clone DKFZp434B094
 C:Genetics:
 A:Note: DKFZp434B094.1
 C:Superfamily: bromodomain homology
 F:113-168/Domain: bromodomain homology <BRO>

alignment_scores:
 Quality: 255.50 Length: 369
 Ratio: 1.412 Gaps: 12
 Percent Similarity: 49.051 Percent Identity: 24.932

align seg 1/1 to: T12534 from: 1 to: 715

```

302 CCTCTCGAAGACAAAAACATGATGACAAACAGAGACGAAAGCGGA 351
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 ProleuleuArgLeuGlnSer.SerLeuGlnSerGlnArgSerSerG 33
352 AAAAGACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 InGlnATGGLuAsnAspGlnGluMetLysAlaAlaLysGluLysL 49
402 AAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 TyrTrpGlnArgLeuArgHisAspLeuGlnArgLysLeuLeuLeu 66
435 .....CGAGATCGAGACCGGGTGGAGATGAGGAGAGAGAGATCTCC 477
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 uLeuLeuArgLysArgGlnLysLysLysArgGlnLysValGlnG 83
478 AGTGTACGACCGCCCTGTGATTAAGCTGCTCCGAGAGAGCTCTGACA 527
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 InValAlaMetGlnLeuArgLeu..... 90
528 AGCTCTTTAGCCAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
91 ..... 96
578 TTTGATCACTGATGACAGATTCGACAGAAAGATCCAGTCTTCT 627
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96 uLeuArgSerValLeuAspGlnLeuGlnAspLysAspProAlaArgL 113
628 TTTGATTTTCCTGTCGATGATTTTATGCTCCGCGTACCTCATGATCTT 677
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 heAlaGlnProValSerLeuLysGluValProAspTyrLeuAspHis 129
678 AAACACCCATGATTTAGTACATGAAAGAAAGATCAAGACAAATGA 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 LysHisPrometAspPheAlaThrMetArgLysArgLeuGlnAlaGln 146
728 CTATAGTCACATAGAAAGAAAGATTAAGTACCTCAACATGATGATG 777
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 yTyrLysAsnLeuHisGlnPheGlnGluAspPheAspLeuLeuLeu 163
778 ATGCCATGATTTAGTAAACAGACACATTTATTAAGCTGCAAG 827
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 snCysMetLysTyrAsnAlaArgAspThrValPheTyrArgLysAla 179
828 AAGCTGTGCTCACTCAAGAAATTTCTTACCGAGAAAGATTCAGAG 877
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 ArgLeuArgAspGlnGlnGlnValValLeuArgGlnAlaArg..... 193
    
```

```

878 CCTGAGCAGAGCATAGATTCATGGCTGACTTGACAGAAACCTCGAAGC 927
193 ..... 193
928 AGAAAGATGAGACAGACACCTCACAGAGTGGGAGACGAGGCTGCTGG 977
193 ..... 193
978 CAGAGAGAGAGAGAGACTCTGGAGATGCCGAGCA.....CAGGC 1018
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 ...ArgGlnValAspSerIleGlyLeuGlnGlnValAspSerLysLeu 209
1019 CTTCAGACGTCGCCAGCAAGAAATATAAAG.....AAAGACA 1056
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 uProGlnArgProAlaAlaAlaProArgArgProPheSerTrrpGln 226
1057 AAGATATGCTTGAAGATTAAGTTAAAGCAATATTA.....GAG 1097
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 aLAspArgLeuLeuAspProAlaAsnArgAlaHisLeuGlnGlnGlu 242
1098 AGAGAGCAGAGACAGCTTGACCCGATC.....GTGAAAGA 1132
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 GlnLeuArgGlnLeuLeuAspMetLeuAspLeuThrCysAlaMetLys 259
1133 ATCTGAGAGAAAGCTGACACAG...CGGCTTGAGACAGTCAAGCGCAT 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 rSerGlySerTrpSerLysArgAlaLysLeuLeuLysLysGlnLeuAla 276
1180 TTGAAGAGAAAGAAACAGATGAGAAACAGAGCTGGGAGCTTCTCATCT 1229
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 euleuArgAsnLys.....LeuSerGlnGlnHisSer 286
1230 GTGATCCCATTTGAGAGAGACCGACTACTGCTGTGAGACTGGGAAT 1279
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 ...GlnProLeuProThrGlyProGly..... 294
1280 GACAACTGGAAGACTTGAGTGTGAGATGATTAAGTTCAGGAGGTTCAAG 1329
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 .....LeuGlnGlyPheGlnG 300
1330 AGGAT 1334
|||||
300 LuAsp 301
    
```

seq_name: p1r2:JC2069

seq_documentation_block:

zinc-finger protein, BR140 - human
 N:Alternate names: Bromodomain protein
 C:Species: Homo sapiens (man)
 C>Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 12-Sep-1997
 C:Accession: JC2069

R:Thompson, K.A.; Wang, B.; Argaves, W.S.; Glancottl, F.G.; Schranck, D.P.; Ruoslahti
 Biochem. Biophys. Res. Commun. 198, 1143-1152, 1994
 A:Title: BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TF
 A:Reference number: JC2069; MUID:94151726

A:Accession: JC2069
 A:Molecule type: mRNA
 A:Residues: 1-1214 <THO>

A:Cross-references: GB:M91585
 C:Comment: This is a nuclear protein with broad tissue distribution, but is especially
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 C:Keywords: DNA binding; phosphoprotein; transcription regulation; zinc finger

F:653-708/Domain: bromodomain homology <BRO>
 F:23,28,41,44/Binding site: zinc (Cys, Cys, His, His) #status predicted
 F:120,205,462/Binding site: phosphate (Ser) (covalent) (by casein Kinase II) #status
 F:276,279,293,296/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
 F:301,304,317,320/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
 F:330,333,350,353/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
 F:386,389,401,405/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
 F:410,413,444,447/Binding site: zinc (His, Cys, Cys, His) #status predicted

Ratio: 0.635 Gaps: 42
Percent similarity: 41.744 Percent identity: 21.279

Alignment block:
US-09-687-230-1 x S39162 ..

Align seg 1/1 to: S39162 from: 1 to: 2440

```

17 CCGCCGCGGCGCGCGCC...CCTGCCCTGCGGCGGCGGCTCGCG 63
   |||||
874 ProProGlyMetThrProProGlnProAla..... 883
64 GCGCCGCGCTCCGCGCTCGCTGCGCTGCGCGGACCGGA...AGCGCG 110
   |||||
884 AlaProThrGlnProSerThrProValSerSerSerGlyGlnProT 900
111 CGCGACGCGCTGGG...CCTGGGCGGCGGCGGCGGCGGCGCG 154
   |||||
900 hProThrProGlySerValProSerAlaThrGlnThrGlnSerThrPro 916
155 GTCGGACATGGGCAAGACACAGACAGCAAGTC.....GGACA 195
   |||||
917 Thr.....ValGlnAlaAlaAlaGlnAlaGlnValThrProGlnPro 931
196 AACACCTCTACGAGAGATATGTAGAGAGACCTTGAAGCTGCTCTCA 245
   |||||
931 nThrProVal.....GlnProProSerValAlaThrProGlnS 944
246 GTAGGAGGAGACGAGT.....CACGAACTCTCCAGCG... 280
   |||||
944 eSerGlnGlnGlnProThrProValHisAlaGlnProProGlnThrPro 960
281 .....CAGCTCGGGGACAGCTCCAG.....CCTCTTCAAGACA 315
   |||||
961 LeuSerGlnAlaAlaSerIleAspAsnArgValProThrProSerThr 977
316 AAAACGATATGCAACACACAGACAG..... 343
   |||||
977 rValAlaSerAlaGlnThrAsnSerGlnInProGlyProAspValPro 994
344 .....AAGCGAAGAAAGAGAAAGAGAGAGAGACAGATTC 382
   |||||
994 aLeuGlnMetCysThrGlnThrGlnAlaGlnAspThrGlnProAsp.Pr 1010
383 AGGGAGAGAAAGGAGAGAAAGAGAGAGAGATTAAGAGAT..... 425
   |||||
1010 oGlyGlnSerIleGlyGlnProArgSerGlnMetGlnGlnAspLeuG 1027
426 .....AAAAAGACGACATCGACACCGCTGACAGATGACGA 464
   |||||
1027 lnglYAlaSerGlnValIleGlyGlnThrAspIleAlaGlnGlnIle 1043
465 GAAAGAGATCTCCAGTGTACGCGCCCTGAGATTAAGATCTGCTCTCA 514
   |||||
1044 GluPro.....MetGluValGlnAspIle 1051
515 GAAGCTCTCACAGCTCTTGAACAAGAGAGAGAGATGAGACAGA. 563
   |||||
1051 sIleProGlnValIleValGlnValIleGlnGlnGlnSerSerSera 1068
563 ..... 563
1068 snGlyThrAlaSerGlnSerThrSerProSerGlnProArgIleGlyS 1084
564 .....CCCTTCAGAGAGCTTGAATCAACTGATGAGACAAAT 601
   |||||
1085 PheIleProGlnGlnGlnLeuArgGlnAlaLeuMetProThrLeuGlnAla 1101
602 GCAGAGAGAGAGATCCAGTGTTC...TTTCATTTCTGTGACTGATT 648
   |||||
1101 uTyIleArgGlnAspProGlnSerLeuProPheArgGlnProValAsp 1118
649 TTATTT.....GCTCGTGGCTACTGCATGATCATTTAAACACCA 692

```

```

1118 InLeuLeuGlyIleProAspTyIlePheAspIleValIleAsnIle 1134
693 TTATGATCATGAGAGAGAGATCAAGACAAATGATCTGATCATCATG 742
   |||||
1135 LeuSerThrIleIleValArgIleValAspThrGlnGlnIleProThr 1151
743 AGAAGTAAAGATTAATCTCAACTAATGTATGATTAATGATGATTA 792
   |||||
1151 pGlnIleValAlaAspAspValThrPheMetIlePheAsnAlaThrLeu 1168
793 ATAAACAGAGACCATTTATTAAGCTGCAAGAGAGATCTGTGAC 842
   |||||
1168 snArgIleThrSerArgValTyIleValPheCysSerIleLeu..... 1181
843 GGAATGAAGAAATCTTACGACAGAGAAAGAAATTCAGACCTGAAG 892
   |||||
1182 ..AlaGlnValPheGlnGlnIleValIleAspProValMetGlnSer 1196
893 AGACTTCATGGCT..... 905
   |||||
1196 uGlyTyIleCysCysGlyArgIleTyIlePheSerProGlnThrLeu 1213
906 .....GACTTCGAGAAACTCGAAGAGAGAGATGAGAGACACCTCA 950
   |||||
1213 ySTyIleValGlnIleValCysThrIleProArgAspAlaIleTyIle 1229
951 CAGAGTGGGAGAGACGAGAGCTGC.....TGCGAGAG 982
   |||||
1230 TyIleGlnAsnArgTyIleHisPheCysGlnIleCysCysPheThr 1246
983 AGAGAGAGAGAGACTCTGAGATGCCGAAAGACACACCGCTTCAAG 1032
   |||||
1246 yGlnAsnValThrLeuGlnAsp.....AspProS 1256
1033 GCAAGAGAAATTAAGAGAGAGAGAGAGATCTGAGATTAAGTTAA 1082
   |||||
1256 eGlnProGlnThrThrIleSerIleAspGlnPheGlnIleCys... 1271
1083 AGCAATTAATTTAGAG.....AG 1099
   |||||
1272 AsnAspThrLeuAspProGlnProPheValAspCysIleGlyArg 1288
1100 AGAGCAGAGACAGCTT..... 1115
   |||||
1288 yIleMetHisGlnIleCysValLeuHisIleTyIleAspIleThrPro 1305
1116 .....GACCGCATCGTGAAGAGATCGA..... 1139
1305 lYpHeValIleCysAspAsnCysIleValIleGlnIleArgProArg 1321
1140 .....GGAAAGCTGACCGAGCGCT 1159
   |||||
1322 AsnIlePheSerAlaIleValArgLeuGlnThrThrArgLeuGlnAla 1338
1160 TGTGAACAGCTAGTCCGAATTTGAAGAGAGAGAGAGATGAGACAGA 1209
   |||||
1338 uIleAspArgValAsnIleValPheLeuArgArgGlnAsn..... 1350
1210 CTTGGAGACTCTCCATCTCGTGGATCCCATGTGTAGAGAGAGAGAG 1259
   |||||
1351 .....HisPro.....GlnAlaGlnGln 1356
1260 TGCCTGTGAGAGCTGGAGATGAGAGAGAGACTGAGTGTGAGTGA 1309
   |||||
1357 ValPheValArgValValAlaSerSer..... 1365
1310 TACTTTGACGGGTTCAAGAGAGATTAAGAGAGAGAGAGATCTCA 1359
   |||||
1366 .....AspIleThrValIleValIleValIleValIleValIle 1375
1360 TA.....TATTTGAATTAATGGCGCTACAGTTCTTATACACCGCAT 1400
   |||||

```

```

1375 eLysSerArgPheValAspSerGlyGluMetSerGlySerPheProTyr 1391
1401 .....TATGACTCCAC 1411
1392 ArgThrLysAlaLeuPheAlaPheGluGluIleAspGlyValAspValGly 1408
1412 ATTTGCAATATTCAGCAGAGATGTTGATTAAATC..... 1448
1408 sPhePheGlyMetHisValGlnAspThrAlaLeuIleAlaProHisGlnI 1425
1449 .....TATCAACCTATGGGAAACACTGATCTCCAACT 1484
1425 IeGlnIleGlyValTyrIleSerTyrLeuAsp..... 1435
1485 GATTTCAGCATCATGATTTTGGCCACCTGC..... 1517
1436 .....SerIleHisPheArgProArgCysLeuArgThrAlaValTyr 1450
1518 .....CAGATTATCCGATG 1533
1450 rHisGluIleLeuIleGlyTyrLeuGluTyrValLysLysLeuValTyr 1467
1534 TCATGGCAGATAGTTTACTGATGTTTAAACAAAGAGG..... 1574
1467 alThrAlaHisIleTrrPAlaCysProProSerGluGlyAspPrrTyrIle 1483
1575 .....CATTCAGAGCCCTACA 1591
1484 PheHisCysHisProProAspGlnLysIleProLysProLysArgLeuG 1500
1592 AGAGATGGAGATGTCATTCGCTGAAGATGAAGCCATACATAGACACTTG 1641
1500 nGluTrrTyrLysLysMetLeuAspLysAlaPheAlaGluArgIleLea 1517
1642 ACACAGAGAAAGAAATGAGACAGATTCACAGATGAGACCCAGGCGGT 1691
1517 sHisPrrTrrLysAspIle..... 1522
1692 TTGGACTCCAGTACTCAAGACAGGCTCATAGCCCTGAAGCAGTAAACA 1741
1523 PheLysGlnAlaAsnGlnAspArgLeuThrSerAlaLysGluLeuProTy 1539
1742 TTTGGCGGTCCAGTGAAGTTTGGACTCGAAGAGCTGAATATTCC 1791
1539 rPhe.....GluGlyAspPheTrrP 1546
1792 AGAGAAACTTGATGAGACACACAGATTCCTCAGAGAACTCAGAGAAC 1841
1546 roAsnValLeuGluGluSer.....IleLysGluLeuGluGlu 1559
1842 CAGATGAAGCT.....TTGAGCACACAGACCC 1870
1560 GluGluGlnArgLysLysGluGluSerThrAlaAlaSerGluTrrProG 1576
1871 TGGGAACATGATCTGTCTTGGTCCCTCATCAGAGAAATGATCTTGC 1920
1576 uGlySerIle.....GlyAspSer..... 1582
1921 TGACAACTGACCAATATCTTAAAGATTCACAGCAGAGTAACTCAGG 1970
1583 .....LysAsnAlaLysLysLysAsnAsnLys 1591
1971 TGATATCTGAAGCAGCTATGAGT...TCGAAAGCAATGGGATTTTCCA 2017
1592 LysThrAsnLysAsnLysSerSerIleSerArgAlaAsnLysLysLysP 1608
2018 TTCCTTCCCGCTCATGGAAAAACAATT 2045
1608 oSerMetProAsnValSerAsnAspLeu 1617
seq_name: pliz:S39161
seq_documentation_block:

```

CREB-binding protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 20-Apr-2000
C:Accession: S39161
R:Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Hagiwara, M.; Montminy, M.R.; Goodman, R.H.
Nature 365, 855-859, 1993
A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.
A:Reference number: S39161; MUID:94019866
A:Accession: S39161
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2441 <CHR>
A:Cross-references: GB:S66385; NID:9435854; PUDN:AB28651.1; PID:9435855
A:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1112-1169/Domain: bromodomain homology <BRO>

```

alignment_scores:
  Quality: 227.50      Length: 894
  Ratio: 0.625         Gaps: 42
  Percent Similarity: 40.716      Percent Identity: 20.470

alignment_block:
US-09-687-230-1 x S39161 ..

Align seg 1/1 to: S39161 from: 1 to: 2441

11 ATCGGCGCGCGCGCGCGCGCGCG.....CCCGCTCGCGCGCGCG 54
   :::::  :::::  :::::  :::::  :::::
838 LeuAlaProGlnAlaSerGlnLeuProCysProProValThrGlnSerPr 854
55 GGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG..... 97
   :::::  :::::  :::::  :::::  :::::
854 OleuHis.....ProThrProProAlaSerThrAlaLacIlyMetP 869
98 .....ACGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 130
   :::::  :::::  :::::  :::::  :::::
869 roSerLeuGlnHisProThrAlaProGlyMetThrProProGlnProAla 885
131 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG..... 166
   :::::  :::::  :::::  :::::  :::::
886 AlaProThrGlnProSerThrProValSerSerGlyGlnThrProThrPr 902
166 ..... 166
902 oThrProGlySerValProSerAlaAlaGlnThrGlnSerThrProThrV 919
167 ..CAAGAAGCACAAAGACACAAAGTC.....GGACAAACACCTCTA 205
   :::::  :::::  :::::  :::::  :::::
919 alGlnAlaAlaGlnAlaGlnAlaGlnValThrProGlnProGlnThrProVal 935
206 CGAGGAGTATGTAGAAAGCCCTGAAGCTGTCTCTCAAACTAGAGAGGA 255
   :::::  :::::  :::::  :::::  :::::
936 .....GlnProProSerValAlaThrProGlnSerSerGlnI 948
256 ACGAAGT.....CACGAACCTCTCCACGG.....CAGC 284
   :::::  :::::  :::::  :::::  :::::
948 nGlnProThrProValHisThrGlnProProGlyThrProLeuSerGlnA 965
285 TCGGGGACAGACTCCAG.....CCTCTCGAAGCAAAAACGATCA 325
   :::::  :::::  :::::  :::::  :::::
965 laAlaIleSerIleAspAsnArgValProThrProSerThrValThrSer 981
326 TGACAAACACAAAGACAG..... 343
   :::::  :::::  :::::  :::::  :::::
982 AlaGluThrSerSerGlnGlnProGlyProAspValPrometLeuGluMe 998
344 ..AAACGGAAAAAGAGAAAGAGAGAGACAGATTCCAGGGGAGAGAA 392
   :::::  :::::  :::::  :::::  :::::
998 tLysThrGluValGlnThrAspAlaGluProGlu.ProThrGluSer 1014
393 AAGGGAGAAAACGGAGAAAGCTTAAGAGAT..... 425
   :::::  :::::  :::::  :::::  :::::

```



```

1790 CCAGAGAACTGATGAGCACCACAGATTGCTCAGGAACTCCAGAG 1839
      :      |||:|||||:      :      :      :      :
1546 pProxValLeuGlucLuser.....IleYsGluLeuGlucL 1560
      :      :      :      :      :      :      :
1840 CCCAGATGAAGCT.....TTGAGCACCAGACCC 1868
      :      :      :      :      :      :      :
1560 LucLugLugLuarGlyLysGluGluSerThrAlaIaSerGluThrPro 1576
      :      :      :      :      :      :      :
1869 CCGGGAACATGATCTGCTCTTGGCTCCCTCATCAGAGAAATGATCTT 1918
      |||:|||||:      :      :      :      :
1577 GluGlySerGln.....GlyAspSer..... 1583
      :      :      :      :      :      :      :
1919 GCGAGAACATGACACATATCTTAAGAAATGACAGACAAATACCCA 1968
      |||:|||||:      :      :      :      :
1584 .....LysAsnAlaLysLysLysAsnAsnL 1592
      :      :      :      :      :      :      :
1969 GGTGATATCTGAACACGTATGAGT...TCGAAAGCAATGGGATTTTC 2015
      :      :      :      :      :      :      :
1592 yAlaSerThrAsnLysAsnLysSerSerIleSerArgAlaAsnLysLys 1608
      :      :      :      :      :      :      :
2016 CATTCCTCCCCCGCTCATGGAAACAACACTT 2045
      |||:|||||:      :      :      :      :
1609 ProSerMetProAsnValSerAsnAspLeu 1618

seq_name: pIR2:A47371

seq_documentation_block:
transcription initiation factor IID 230K chain - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C.Accession: A47371
R.Kokubo, T.; Gong, D.W.; Yamashita, S.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Genes Dev. 7, 1033-1046, 1993
A.Title: Drosophila 230-kb TFIID subunit, a functional homolog of the human cell cycle
A.Reference number: A47371; MUID:93279463
A.Accession: A47371
A.Status: preliminary
A.Molecule type: mRNA; protein
A.Residues: 1-2068 <KOK>
A.Cross-references: GB:S61883; NID:g385550; PIDN:AA826991.1; PID:g385551
A.Note: sequence inconsistent with nucleotide translation
A.Note: sequence extracted from NCBI backbone (NCBIN:133002, NCBIPI:133003)
C.Genetics:
A.Gene: FlyBase:Ta1250
A.Cross-references: FlyBase:FBgn0010355
C.Superfamily: unassigned bromodomain proteins; bromodomain homology
C.Keywords: transcription initiation
F:1498-1553/Domain: bromodomain homology <BRO1>
F:1620-1675/Domain: bromodomain homology <BRO>

alignment_scores:
      Quality: 217.50      Length: 701
      Ratio: 0.732      Gaps: 31
      Percent Similarity: 42.368      Percent Identity: 19.829

alignment_block:
US-09-687-230-1 x A47371 ..
Align seg 1/1 to: A47371 from: 1 to: 2068

99 CCGAGAACGCGCGCGACGCGCTGGCGCGGCGGCGGCGGCGGCGGCGG 148
   ||| :|| :||| :||| :||| :||| :||| :||| :||| :|||
1306 ProGluYMetProThrSerLeuGlyAspProLysSerSerGly..... 1320
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
149 GCGCGCGGCGGACATGGGCAAGAGACACACACACATGCGGACAAAC 198
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1321 .....GlyHisSerHisLysGluArgAspSerGly.... 1330
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
199 ACCTTACGAGAGAG.....TATGTAGAGAGAGCC 227
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1331 .....TyrLysGluValSerProSerArgLysLysPheLysLeuLysPhe 1345

```

```

228 TTGAAGTGTGCTCCTCAAAAGTAGAGGG...AAGAAAGTCACCAACTCTC 274
      ||| ||||| ||||| :||| :||| :||| :||| :||| :|||
1346 ...AspLeuLysLeuLysCysGlyAlaCysGlyLValGlyHisMetArg 1361
      :      :      :      :      :      :      :
275 CACGGCGCAGC.....TCGGGGCAGCAGCTCCAGCTCTTCG 309
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1361 gThrAsnLysAlaCysProLeuTyrSerGlyMetGlnSerSerLeuSerG 1378
      :      :      :      :      :      :      :
310 AAGACAAA.....AACGATCATGACAAACAAAGACAGACA... 344
      :      :      :      :      :      :      :
1378 InSerAsnProSerLeuAlaAspAspPheAspGlnSerGluLysGlu 1394
      :      :      :      :      :      :      :
344 ..... 344
      :      :      :      :      :      :      :
1395 MetThrMetAspAspAspLeuValAsnValAspGlyThrLysValThr 1411
      :      :      :      :      :      :      :
345 .....AAGCGAAA.....AAGCAAA 360
      :      :      :      :      :      :      :
1411 rLeuSerSerLysAlaLeuLysArgHisGlyAspAspGlyLysArg 1428
      :      :      :      :      :      :      :
361 AGAAGAGAGAGAG.....CAGATTCAGGGGAGAAAGAAAG 395
      :      :      :      :      :      :      :
1428 rGSerGlySerSerSerGlyPheThrLeuLysValProArgAspAlaMet 1444
      :      :      :      :      :      :      :
396 GGGAGAAAGAGAGAGAGATTAAAGAGATAAAGAGAGAGATCGAGA 445
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1445 GlyLysLysLysArgVal..... 1451
      :      :      :      :      :      :      :
446 CCGGGTGAAGAAATGAGCGACAAAGAAAGATCTCAGCTGACGCCCTGCA 495
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1452 .....GlyGlyAspLeuHisCysAspTyrLeuGlnA 1462
      :      :      :      :      :      :      :
496 GATTAGACTTGGCTCTGAGAGCGCTTCACAAAGCTTTAGCCAAACAA 545
      || :|| :||| :||| :||| :||| :||| :||| :||| :|||
1462 rGHisAsnLysThrAlaAsnArgArgThrAspProVal..... 1475
      :      :      :      :      :      :      :
546 GAAGAAAGTAGAACAGACACCCCTCAAGAAAGCTTGAATCAAGTAGAG 595
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1476 .....ValValLeuSerSerIleLeuGlnLleHisAsnGlnLeuArg 1490
      :      :      :      :      :      :      :
596 ACAATTCACAGAAAGATCCAAAGCTCTTTCATTTACCTTCGTAAGT 645
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1490 gSerMetProAspValSerPro.....PheLeuPheProValSerA 1504
      :      :      :      :      :      :      :
646 ATTTTATGCTCCTGGCTACTCCATGATCATTAACACCCCAATGGATTT 695
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1504 lAlaLysValProAspTyrTyrArgValAlaThrLysProMetAspLeu 1520
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
696 AGTACCATGAAGAAAGATCAAGACATGACATGACATGCCATCAAGAGA 745
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1521 GlnThrMetArgGlnTyrTyrLeuArgGlnArgTyrTyrThrSerArgLys 1537
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
746 ACTAAAGATTAATCTCAAACTAATGTGTACTAATGCCATGATTTACATA 795
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1537 tPheLeuGlnAspLeuLysGlnIleValAspAsnSerLeuIleTyrAsnG 1554
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
796 AACCAAGACCAATTTATTAAGCTCAAGAAAGCTGTGCACATCAAGA 845
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1554 lProGlnSerAlaTyrThrLeuAlaAlaGlnArgMetPheSerSerCys 1570
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
846 ATGAATAATCTTAC.....CAGGAAGAAATTGAGAGCCTGGAAGAGAG 889
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1571 PheIleLeuLeuAlaGluArgGlnAspLysLeuMetArgLeuGlnLysAl 1587
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
890 CATAGACTTCATGGCTGAC..... 908
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1587 aIleAsnProLeuLeuSerAspAspArgGlnValAlaLeuSerPheIle 1604
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
909 .....TTCAGAAAACTCGAAAGACAGAAAGATGGAACAGACACCTCA 950
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1604 heAspLysLeuGlnHisSerGlnIleLysGlnLeu..... 1614
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
951 CAGAGTGGGAGAGAGAGAGCTGCTGGCAGAGAGAGAGAGAGAGACTGCG 1000

```

```

1614 ..... 1614
1001 AGATGCCGAGACACGCTTCAGAGTCCAGCAAGAAATATAAAGA 1050
1615 ....ProGluSerTTPProPheLeuLysPro.....ValAsnLysLysG 1628
1051 AAGACAAGATATG.....CTTGAAGAT 1073
1628 InValLysAspTyrThrValLLeLysArgProMetAspLeuGluThr 1644
1074 AAGTTAAAGCAATATTTAAGAGAGACGAGCAGCTTGACCGCAT 1123
1645 lLeGlyLysAsnLLeGluAlaHisArgTyrHisSerArgAlaGluTyrLe 1661
1124 CGTGAAGGATCTGGAGGAAAGCTGACCGGCTTGGAACAGTCAGT 1173
1661 uAla.....AspLLeGluLeuLLeAlaThrAsnC 1671
1174 GCGAA.....TTTGAAGAGA 1190
1671 ySGLuGlnTyrAsnGlySerAspThrArgTyrThrLysPheSerLysLys 1687
1191 AAACCAAGATGGAACACACGCTGGGACTTCTCCATCCTGTGATCCAT 1240
1688 lLeuGluTyrAlaGlnThrGln..... 1695
1241 TGTAGAGAGCCAGGCTACTGCTGAGACTGGGAATGACACTGGA 1290
1696 .....LeuLLeGluPheSerGlnHisCysGlyG 1705
1291 GACTTCAGCTGGAGTGAATCTTGCAGGGGCTCAAGAGATAAAGC 1340
1705 lInLeGluAsnAsnLLeAlaLysThrGln.....GluArgAlaArg 1718
1341 AACAAAGTACCTCCAGCTGTATATTTGAATATGGCCCTACAGTCTTA 1390
1719 GluAsn..... 1720
1391 TGCACCGCATTAATGATCCACATTTGCAAAATATCAGAAAGATGATCTG 1440
1721 .AlaProGluPheAspGlnAlaTPrGlyAsn.....AspArgTyrA 1734
1441 ATTATATGATTAATCAACSTATGGGGAAGACSTGATCTTCCAAAGATATTC 1490
1734 snPhe.....AspArgLysSerArgAlaSerSerProGlyLysP... 1746
1491 AGCATCCATGAGTTTGGCCACGTCGCAAGATTTCCGTAATGATGCG 1540
1747 .....AspTyr..... 1748
1541 AGATAGTTACTGATGTTTAAACAAAGAGAGGATCCAGACCCCTAC 1590
1749 .....lLeAspValGluGlnHisGlyLysHisAlaSerSerSera 1762
1591 AAGAGATGAGATGTCATGCTGCTGAGATGAAGCCATAGTACACTT 1640
1762 snSerLLeHisArgSerMetGlyAlaGlnAlaGlySerSerHisThrAla 1778
1641 GACACAGGAAAGAAATGAGACGATTAACAGAAATGAGACCCAGG... 1688
1779 ProAla.....ValArgLysProAlaProProGlyPyr 1789
1688 ..... 1688
1789 GGLyGluValLysArgGlyArgGlyArgProAlaGlyGlnArgAspProV 1806
1689 .....CGTTTGACTTCAGTACTCAAGACAGGCTCATAGCCGTGAAA 1730
1806 aLGlGluLAspLeuGlnCysSerThrAspAspGln..... 1817
1731 GCAGTAACAATTTTGGCGTCCAGTTGAGTTTTGGACTGTCGAGAAAGC 1780

```

```

1818 .....AspAspAspLLeuG 1822
1781 TGAATATTCAGAGAAACTTGATGAGACCAACCAATTTGTCAGAGAC 1830
1822 uGluAspPheGln.....GluV 1828
1831 TCCAGGAGCCGAGAAATGAACTTTGAGCAGACAGACCCCTGGAAATG 1880
1828 aISerClnAspLLeuAsnAlaAlaSerLLeuAspGlnGlyGluArg 1844
1881 ATC 1883
1845 lLe 1845
seq_name: p1r2:A54277
seq_documentation_block:
transcription adaptor protein p300 - human
C/Species: Homo sapiens (man)
C/Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 10-Dec-1999
C/Accession: A54277; S60344
R/Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; Decaprio, J.A.; Lawrence, J.B.; Li
Genes Dev. 8, 869-884, 1994
A/Title: Molecular cloning and functional analysis of the adenovirus E1A-associated 3
A/Reference number: A54277; MUID:95011587
A/Accession: A54277
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2414 <ECK>
A/Cross-references: GB:001877; NID:g495300; PIDN:AAA18639.1; PID:g495301
A/Note: In the authors' translation 941-Ser is shown after 961 and consequently, res1
R/Lundblad, J.R.; Kwok, R.P.S.; Laurance, M.E.; Harter, M.L.; Goodman, R.H.
Nature 374, 85-88, 1995
A/Title: Adenoviral E1A-associated protein p300 as a functional homologue of the tran
A/Reference number: S60344; MUID:95174889
A/Accession: S60344
A/Status: preliminary
A/Molecule type: protein
A/Residues: 552-660 <LDN>
C/Genetics:
A/Gene: GDB:EP300
A/Cross-references: GDB:9862958; OMIM:502700
A/Map position: 22q13.2-22q13.2
C/Superfamily: unassigned bromodomain proteins; bromodomain homology
C/Keywords: phosphoprotein; transcription; zinc finger
F:1075-1132/Domain: bromodomain homology <RBO>
F:89,507,1136,1295,1497,1834,1977,2062,2320/Binding site: phosphate (Ser) (covalent)
F:1734/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predict
alignment_scores:
Quality: 216.00 Length: 740
Ratio: 0.663 Gaps: 33
Percent Similarity: 44.054 Percent Identity: 21.216
alignment_block:
US-09-687-230-1 x A54277 ..
Align seg 1/1 to: A54277 from: 1 to: 2414
17 CCGCCCGGC.....GCGCGCGCCCTGCTGCTGCGCGCGG 54
|||||
870 ProProGlyProGlnSerGlnAlaLeuHisProPro..... 882
55 GGTCTCGCGCGCGCGCTCCGCTCGCTGCTGCGCGCGGAG 104
|||||
883 ....ArgGlnThrProThrProThrThrGlnLeuProGlnValG 898
105 GCGCGCGCGCACGCGCTGCGCGCGCGCGCGCGCGCGCG 154
:::
898 lInProSerLeuProAlaAlaProSerAlaAspGlnProGlnGlnInPro 914
155 GTCGACATGGCGCAAGAACACAGAGACCAAGTC..... 190

```

915ArgSerGlnGlnSerThrAlaAlaSerValProThrProAsnAl 929
191GACAAACACCTTCAGA.....GGATAT 215
929 AProLeuLeuProGlnProAlaThrProLeuSerGlnProAlaValS 946
216 GTAGAGAAAGCCCTTGAAAGCTGCTCCAAAGTAGAGGGAACGAAGTCAC 265
946 erileGlnGlnValSerAsnProProSer.....ThrSerSer 959
266 CGAACTCTCAGGGGAGCTCGGGGAGAGCTCCAGCTCTTCGAACACA 315
960 ThrGlnValSerGlnAlaIleAlaGlnGlnPro..... 972
316 AAAACGATCATGACAAACAGAGACGAAAGGAAAGGAAAGGAAAGAAA 365
973SerGlnGlnValLysMetGlnAlaLysMetGlnVal 984
366 GGAGAGAGAGATTCAGGGGGAAGAAAGGGAAGAAACGAGAGAGAT 415
984 AlaSerGlnProGlnProAlaAspThrGlnProGlnAspIleSerGlnUser 1000
416 TAAGAGAGATTAAGAAAGCAGATCCAGACCGGGTGGAGAATGAGCAG 465
1001 .LysValGlnAspCysLysMetGlnSerThrGlnThrGln...GlnLys 1016
466 AAAAAGATCTCAGTGCAGCCCGCTGAGATTAAGATTCGCTCCTCAG 515
1016 erThrGlnLeuLysThrGlnIleLysGlnGlnGlnLysGlnProSerThr 1032
516 AAGCCTCTCAGACGCTTTAGCC.....AAACAAGAAAGAT 553
1033 SerAlaThrGlnSerSerProAlaProGlnLysLysLysLysLysLys 1049
554 AGAACAAGACCCCTCAGAAGCTTGAAATCACTGATGAGAGACAATGC 603
1049 eLysProGlnLysLeuArgGlnAlaLeuMetProThrLeuGlnAlaLeu 1066
604 AGAACAAGATCAAGATGCTTTC...TTTTCATTTTCTGATCACTCATTT 650
1066 yThrGlnAspProGlnSerLeuProPheArgGlnProValAspProGln 1082
651 ATT.....GCTCTGCTACTCCATGATCATTAACACCCCAATGATTT 694
1083 LeuLeuGlnIleProAspTyrPheAspIleValLysSerProMetAspLe 1099
695 TAGTACCATGAAGAAAGATCAAGAAATGACTATCATGCTCATATGAAG 744
1099 userThrIleLysArgLysLeuAspThrGlnGlnIleGlnGlnLysPro 1116
745 AACTAAAGATTAATCTCAACTAATGTACTAATGCATGATTTACAT 794
1116 IuTyValAspAspIleThrLeuMetPheAsnAlaIleThrLeuLysAsn 1132
795 AAACAGAGACCATTTATTTAAAGCTCGAAAGAGCTGTTGCATCAG 844
1133 ArgLysThrSerArgValTyrLysTyrCysSerLysLeu..... 1145
845 AATGAATAATCTTACAGAAAGAAATTCAGAGCTGAAGCAGACATAG 894
1146 .SerIuValPheGlnGlnGln...IleAspProValMetGlnSerLeu 1161
895 ACTTCATGCT..... 905
1161 IuTyTyCysGlnGlnArgLysLeuGlnPheSerProGlnThrLeuLysCys 1177
906 ...GACTTCGAGAAACTCGAAAGCAAGATGAGACAGACATCCACA 952
1178 TyrGlnLysGlnLeuLysThrIleProArgAspAlaThrTyTySer 1194
953 GAGTGGGAGAGAGAGCTGC.....TGCAGAGAG 984
.....

1194 rGlnAsnArgTyrHisPheCysGlnLysCysPheAsnGlnIleGlnLys 1211
985 AGAAGAGAGACTGTGAGATGCCGAACACAGCTTTCAAGAGTCCAGC 1034
1211 IuSerValSerLeuGlnAspAspProSerGlnProGlnThrIleAsn 1227
1035 AAAGAAAT...AAAAAGAAAGCAAGATATGCTTGAAGTTAAGTTT... 1079
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGlnLeu 1244
1080AAAAAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115
1244 eValGlnLysThrGlnLysGlnLysArgLysMetHisGlnIleLysValLeu 1261
1116GACCCATCTGAG 1130
1261 IuHisGlnIleIleThrProAlaGlnPheValCysAspGlnCysLeuLys 1277
1131 GAATCTGGA..... 1139
1278 LysSerAlaArgThrArgLysGlnAsnLysPheSerAlaLysArgLeuP 1294
1140GAAAGCTGACAGGCGGCTTGTGAAACAGTCAGTCCG 1176
1294 oSerThrArgLeuGlnIleThrPheLeuGlnAsnArgValAsnAspPheLeu 1311
1177 AATTTGAAAGAAAGAAACCAAGATGAAACACAGCTTGGAGCTTCAT 1226
1311 rGArgGlnAsnHisProGlnSerGlnGlnValThrValArgValValHis 1327
1227 CCTGTGATCCCATTTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
1328 AlaSerAspLysThrValGln.....ValLysProGln 1338
1277 AATGACAGCTGAAAGACTTCAGCTGAGTGAATACCTTCCAGAGGTTCA 1326
1338 yMetLysAlaArgPheValAspSerGln...GlnMetAlaGlnSerPhe. 1353
1327 AAGAGATTAAGAAAGCAAGAGCTCAGCTTGAATTAATTAATGAG 1376
1353 1353
1377 CCCTACAGTCTGTATGACCGGATTAATGATCCATTCGAAATATACAG 1426
1354 ProTyAlaGlnThrLysAla.....LeuPheAlaPheGlnGln 1365
1427 CAAGGATGATTCGATTAATCTAT.....TCAACCTATG 1461
1365 uIleAspGlnValAspLeuLysPhePheGlnyMetHisValGlnGlnLys 1382
1462 GGAAGAGACTGATCTTCCAGATGAT..... 1487
1382 IySerAspCysProProProAsnGlnArgValTyrIleSerTyLeu 1398
1488 TTCACATCATGATGATTTTGGCCAGCTGC.....CAAGATTATCC 1528
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyrHis 1415
1529 GTATGTCATGSCAGATAGTTACTGAGATGTTTAAACAAAGAGAGGAT 1578
1415 sGlnIleLeuIle...GlyTyLeuGlnIuTyValLysLysLeuGlnTy 1431
1579 CCAGAGACCCCTCAAGAGATGAGATGTCATTCCTGAAGATGAAGCCAT 1628
1431 Hr.....ThrGlnHis 1434
1629 ACTAGACACTTGACACAGAAAGAAATGAGACAGATTAAGAAATGAGA 1678
1435 IleThrPalacysProProSerGlnGlnLysAspAspTyrIlePheHisCysHis 1451
1679 GCCACCGAGCGT..... 1691
1451 sProProAspGlnLysIleProLysProLysArgLeuGlnGlnIuThrTyL 1468

```

1692 .....TTGCAC..... 1697
1468 yslYsmelluAspIysAlaValserGIuArgIleValHisAspTyrLys 1484
1698 .....TCCAGTACTCAAGACAGCCTCATACCGCTGAAGACGT 1735
1485 AspIlePheLysGIuAlaThrGIuAspArgLeuThrSerAlaLysGIuLe 1501
1736 AACAAATTTGGCGTTCCAGTTGAAGTTTTCAGCTTGAAAGACGTGAA 1785
1501 uProTyrPhe.....GlucIYAsp 1508
1786 TATTCAGAGAACTTGATGACACCAAGATTGCTCAGGGAATCCAG 1835
1508 heTrpProAsnValLeuGIuGIuSer.....IleLysGIuLeuGIu 1521
1836 GAAGCCAGAAATGACGT 1853
1522 GIuGIuGIuGIuGIuArg 1527

seq_name: pir2:S39580

seq_documentation_block:
HBRM protein - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C:Accession: S39580
R:Muchardt, C.; Yaniv, M.
EMBO J. 12, 4279-4290, 1993
A:Title: A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm gene
A:Reference number: S39580; MUID:94038910
A:Accession: S39580
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1586 <MUC>
A:Cross-references: EMBL:X72889; NID:q414116; PIDN:CAA51407.1; PID:q414117
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:216-249/Region: glutamine-rich
F:1423-1478/Domain: bromodomain homology <BRO>

alignment_scores:
Quality: 210.00 Length: 363
Ratio: 1.235 Gaps: 15
Percent Similarity: 46.832 Percent Identity: 23.967

alignment_block:
US-09-687-230-1 x S39580 ..
Align seg 1/1 to: S39580 from: 1 to: 1586

321 GATCATGACAAACACAGAGAC.....AGAAAGCGAAGAAAGAG 358
||| |||..... ||| |||
1253 AspMetAspArgArgArgGIuAspAlaArgAsnProLysArgLysProAr 1269
359 AAAGAAAGAGAGAGACAGATCCAGGG..... 386
| |||..... |||
1269 gleuMetGIuGIuAspGIuLeuProSerTrpIleIleLysAspAspAlaG 1286
387 .....GAAGAAAG..... 395
||| |||
1286 luValGIuArgLeuThrCysGIuGIuGIuGIuLysIlePheGlyArg 1302
396 ...GGAGAGAAAGAGAGAG..... 410
::: |||..... |||
1303 GlySerArgGIuArgArgAspValAspTyrSerAspAlaLeuThrGIuLys 1319
410 ..... 410

1319 sGIuThrPheuArgAlaIleGIuAspGIuAsnLeuGIuGIuMetGIuGIuG 1336
411 .....AGAGTTAAGAGAGATAAAGAGAGCA.....GATCGAGACCGG 449

```

```

1336 luValArgLeuLysLysArgLysArgArgArgAsnValAspLysAspPro 1352
||| |||..... ||| |||
450 GTGGAGAAATAGACGAGAAAGAAAGATCTCCAGTGCACCCCTGTGAGATT 499
||| |||..... ||| |||
1353 AlaLysAspValGIuLys.....AlaLysLysArgArg 1364
500 AGACTTCCTCCCTGCAGAACCTCTCACAAGCTCT.....TTAGCA 540
||| |||..... ||| |||
1364 gGIuArgProProAlaGIuLysLeuSerProAsnProProLysLeuThrL 1381
541 AACAGAGAA..... 548
||| |||
1381 ysgImetAsnAlaIleIleAspThrCysIleAsnTyrLysAspSerCys 1397
549 GAAGTGAACAGACACCCCTCAAGAAAGCTTG.....AATCA 586
::: |||..... ||| |||
1398 AsnValGIuLysValProSerAsnSerGIuLeuGIuLysAsnSe 1414
587 ACTGATGAGACAATTCGAGAGAAAGATCCAGAGCTTCTTTCATTTG 636
| ||| |||..... ||| |||
1414 rSerGIuArgGIuLeuSer.....GIuValPheIleGIuLeuP 1427
637 CTGTGACTGATTTTATTCTCTCGCTACTCCATGATCATTAACACCA 686
||| |||..... ||| |||
1427 roSerArgLysGIuLeu...ProGIuTyrTrpGIuLeuIleArgLysPro 1442
687 ATGATTTTGTACCATGAAAGAAAGAAAGATCAAGATGATTAATGATG 736
::: ||| |||..... ||| |||
1443 ValAspPheLysLysIleLysGIuArgIleArgAsnHisLysTyrArgSe 1459
737 CATAGAGAACTAAGAGATTAAGTAACTAATGTACTAATGCCATGA 786
||| |||..... ||| |||
1459 rLeuGIuAspLeuGIuLysAspValMetLeuLeuCysHisAsnAlaGIuNT 1476
787 TTTACATATAACCAAGACCATTTATTATAAGCTCCAAAGACGTGTG 836
::: ||| |||..... ||| |||
1476 hrPheAsnLeuGIuGIuSerGIuIleTyrGIuLysSerIleValLeu... 1491
837 CACTCAGAAATGAAATTTCTAGCCAGAGAAAGAAATTCAGAGCTGAAGCA 886
||| |||..... ||| |||
1492 .....GI 1492
887 GAGCATAGACTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
||| |||..... ||| |||
1492 nSerVal...PheLysSerAlaArgGIuLysIleAlaLysGIuGIuLys 1508
937 GAACAGACACCTCAGACAGAGTGGGAGAGGAGAGCCCTGCGGACAGAGAG 986
:: ||| |||..... ||| |||
1508 erGIuAspLeuSerAsnGIuGIuLys.....GIuGIuGIu 1519
987 AGAGAGACTCTGGAGATCCGAAAGCAGACAGCCCTTCMAAGTCCAGCA 1036
||| |||..... ||| |||
1520 AspGIuGIuGIuSerGIuLysSerValLysSerValLysValLysIleY 1536
1037 ACAGAAATTAAGAAAGAAAGACAAAGATATGCTTGACATTAAGTTAAAGC 1085
| ||| |||..... ||| |||
1536 sLeuAsnLysLysAspAspLysGIuArgAspLysGIuLysGIuLysA 1553
1086 .....AATTAATTAGAGAGAGAG 1103
1553 rgProAsnArgGIuLysAlaLysProValValSerAspPheAspSerAsp 1569
1104 CAGAGACAGCTTGACCGCATCTGGAAGAAATCTGGAGAG 1142
::: ||| |||..... ||| |||
1570 GIuGIuGIuAspGIuArgGIuLysGIuSerGIuGIuLysGIuSerGIu 1582

seq_name: pir2:JC5056

seq_documentation_block:
polybrmo 1 - chicken
N:Alternate names: polybromodomain protein 1
C:Species: Gallus gallus (chicken)

```

C.date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 20-Sep-1999
 C.accession: JCS056; S60678
 R.NICOLAS, R.H.; Goodwin, G.H.
 Gene 175, 233-240, 1996
 A.title: Molecular cloning of polybramo, a nuclear protein containing multiple domains 1
 A.reference number: JCS056, MID:97074677
 A.contents: embryo
 A.accession: JCS056
 A.status: nucleic acid sequence not shown
 A.molecule type: mRNA
 A.residues: 1-1633 <NIC>
 A.cross-references: EMBL:X90849; NID:g951230; PID:CAA6235.1; PID:g951231
 A.note: submitted to the EMBL Data Library, August 1995
 C.comment: This is a nuclear protein with five repeats of a domain thought to be involved
 C.genetics:
 A.gene: pbl
 C.superfamily: unassigned bromodomain proteins; bromodomain homology
 F:66-152/Domain: bromodomain repeat <BR1>
 F:73-128/Domain: bromodomain homology <BR01>
 F:198-284/Domain: bromodomain repeat <BR2>
 F:205-260/Domain: bromodomain homology <BR02>
 F:399-485/Domain: bromodomain repeat <BR3>
 F:406-461/Domain: bromodomain homology <BR03>
 F:537-623/Domain: bromodomain repeat <BR4>
 F:544-599/Domain: bromodomain homology <BR04>
 F:675-761/Domain: bromodomain repeat <BR5>
 F:682-737/Domain: bromodomain homology <BR05>

alignment_scores:
 Quality: 207.50 Length: 698
 Ratio: 0.638 Gaps: 33
 Percent Similarity: 46.562 Percent Identity: 21.060

alignment_block:
 us-09-687-230-1 x JCS056 ..

Align seg 1/1 to: JCS056 from: 1 to: 1633

```

144 ACCGGGCGCGTCGACATGGCAGACAGACAGACAGATCGCA 193
      ::::: ||| ||||| |||::: ::::: |||||
499 SerSerLathSerApThrGlySer.....SerLysArgLysSerly 513
      ::::: ||| ||||| |||::: ::::: |||||
194 CAAACACCTCTACGAGAGTATGTAGAGAACCTTGAAGCTGCTCTCA 243
      |||::: |||::: |||::: |||::: |||:::
513 slyasametrlygslargmetlysileutyasnalaValleu. 529
      |||::: |||::: |||::: |||::: |||:::
244 AAGTAGAGGAGACGACGACGACGACGACGACGACGACGACGAC 293
      |||::: |||::: |||::: |||::: |||:::
530 .....GluLalArgLysSerGlyThrGlyArgArgLeucys 541
      |||::: |||::: |||::: |||::: |||:::
294 GACTCAGCCTCTTGAA...GACAAAACGATCAT...GACAAACACAA 337
      |||::: |||::: |||::: |||::: |||:::
542 AspreuPheMetValLysProSerLysLysAspLysProAspLysLys 558
      |||::: |||::: |||::: |||::: |||:::
338 G.....GACAGAAAG.....CGGA 351
      |||::: |||::: |||::: |||::: |||:::
558 silelleucluprometaspLeuLysMetLysLysLysLysLysLys 575
      |||::: |||::: |||::: |||::: |||:::
352 AAAAGAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 395
      |||::: |||::: |||::: |||::: |||:::
575 snarplytYrValglGluLalMetLysAspPheMetLysLysLysLys 591
      |||::: |||::: |||::: |||::: |||:::
396 CGGAGAAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 445
      |||::: |||::: |||::: |||::: |||:::
592 PheArgnaLalArgLysLysLysLysLysLysLysLysLysLysLys 608
      |||::: |||::: |||::: |||::: |||:::
446 C.....CGGTGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
      |||::: |||::: |||::: |||::: |||:::
608 ralnlsmetleuGluLysLysLysLysLysLysLysLysLysLysLys 623
      |||::: |||::: |||::: |||::: |||:::
481 GTCAAGCCCTGTGAGATTAGAC.....TTGCCCTGTGAGAAAGCT 521
      |||::: |||::: |||::: |||::: |||:::

```

```

624 ....GlyProLeuProGluAspAspValAlaSerProLysLeuLys 638
      |||::: |||::: |||::: |||::: |||:::
522 CTCACA.....AGCTCTTAGCCAAACAGAGAGAGAGAGAGAGAG 565
      |||::: |||::: |||::: |||::: |||:::
639 LeuSerArgLysSerGlyLysSerProLysLysSerLysLysMetThr 655
      |||::: |||::: |||::: |||::: |||:::
566 CTTTCAGAAAGCTTTGAATCAACTGATGAGCAATTGCAG..... 605
      |||::: |||::: |||::: |||::: |||:::
655 ometGlnGlnLysLeuAsnGluValTYrGluAlaValLysAsnTYrThra 672
      |||::: |||::: |||::: |||::: |||:::
606 .....AGAAAGATCCAAAGCTCTTCTTCAATTCCGCTGACT 644
      |||::: |||::: |||::: |||::: |||:::
672 splyslArgGlyArgArgLeuSerAlallePheLeuArgLeuProSerArg 688
      |||::: |||::: |||::: |||::: |||:::
645 GATTATTATGCTCCTGCTACTCTCATGATCATTAACACCAATGATTT 694
      |||::: |||::: |||::: |||::: |||:::
689 SerLeuLeu...ProAspTYrTYrLethrLethrLysLysProValAspMe 704
      |||::: |||::: |||::: |||::: |||:::
695 TAGTACCATGAAAGAAAAGATCAAGAACATGACTATGATCCATAGAG 744
      |||::: |||::: |||::: |||::: |||:::
704 tGluLysLleArgSerHisMetMetAlaAsnLysTYrGlnAspLysAsps 721
      |||::: |||::: |||::: |||::: |||:::
745 AACTAAAGAAATTAAGTCAACTATGTTACTATGCGCATGATTACAT 794
      |||::: |||::: |||::: |||::: |||:::
721 ermetValGluAspPheValMetMetPheAsnAlaCysThrTYrAsn 737
      |||::: |||::: |||::: |||::: |||:::
795 AAACCAAGACCATTTATTAAGCTGCA.....AAGAAGCT 832
      |||::: |||::: |||::: |||::: |||:::
738 GluProGluSerLeuLleTYrLysAspAlaLeuValLeuHisLysValLe 754
      |||::: |||::: |||::: |||::: |||:::
833 GTTGACATCA.....G 843
      |||::: |||::: |||::: |||::: |||:::
754 uLeuGluThraArgGluLleGluGluAspGluAspSerHisValProA 771
      |||::: |||::: |||::: |||::: |||:::
844 GAATGAATAATCTTACCCAGAAAGAAATTCAGACGAGAGAGAGATA 893
      |||::: |||::: |||::: |||::: |||:::
771 snValThrLeuLeuLleGlnGlnLeuLleHisLysnLeuPheValSerVal 787
      |||::: |||::: |||::: |||::: |||:::
894 .....GACTTCATGCTGCA 907
      |||::: |||::: |||::: |||::: |||:::
788 MetSerHisGlnAspArgGluLysArgCysTyrSerAspSerLeuAlaGl 804
      |||::: |||::: |||::: |||::: |||:::
908 C..... 908
      |||::: |||::: |||::: |||::: |||:::
804 uLleProAlaValAspProAsnPheProAsnLysProProLeuThrPhea 821
      |||::: |||::: |||::: |||::: |||:::
908 ..... 908
      |||::: |||::: |||::: |||::: |||:::
821 spllelleArgLysAsnValGluAsnAsnArgTYrArgArgLeuAspLeu 837
      |||::: |||::: |||::: |||::: |||:::
909 .....TTGCAGAAAACCTGAAAAGCAGAAAGA 934
      |||::: |||::: |||::: |||::: |||:::
838 PheGlnGluAsnMetPheGluValLeuGluArgLalArgArgMet...As 853
      |||::: |||::: |||::: |||::: |||:::
935 TGGAAACAGACCTTCACAGAGAGTGGGAGAGACGAGCTGCTGGCAGGA 983
      |||::: |||::: |||::: |||::: |||:::
853 nArgThrAsp...SerGluLleTYrGluAspAlaValAlaGluLeuGlnP 869
      |||::: |||::: |||::: |||::: |||:::
984 .....GAGAGAGAGACTGCGAGAT..... 1004
      |||::: |||::: |||::: |||::: |||:::
869 hePheLleLysLleArgAspLleuLysLysAsnGluGluLleLeuLeu 885
      |||::: |||::: |||::: |||::: |||:::
1005 .....GCCGAAGCACACGCTTCAAGATCCACAGCA 1036
      |||::: |||::: |||::: |||::: |||:::
886 SerProAlaLeuSerTYrThrLysHisLysHisAsnAspValGluLys 902
      |||::: |||::: |||::: |||::: |||:::
1037 AGAAATATTAAGAAA...GACAAAGATATGCTTGAAGATTAAGTTAAA 1063
      |||::: |||::: |||::: |||::: |||:::
902 sGluLysLysGluLysLeuProLysGluLleGlnGluAspLysLeuLysA 919
      |||::: |||::: |||::: |||::: |||:::
1084 GCAATTAATTTAGAGAGAGACGAGACGAGCTTGAACCGCATGTGAAGAA 1133
      |||::: |||::: |||::: |||::: |||:::
919 rGluGluGluLysArgGluAlaGluLysSerGluAspSerSerGlySer 935
      |||::: |||::: |||::: |||::: |||:::

```

```

1134 TCTGGAGAAAGCTGACAGGGCGGCTGTGAACGCTCAGTGGCAATTGTA 1183
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
936 AAGGlyLeuSerSerLeuHisAlaThrTyrSerGlnAspSerPheTyr 952
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1184 AAGAAGAAACACAGATGAGACAAAGACAGCTTGGGACCTCTCCATCCCTGTGG 1233
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
952 SAsn.....SerMetTyrHisValGlyAla 960
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1234 ATCCCATTTGTAGAGAGCCAGGCTACTGCTGTGAGACATGGGAATGACA 1283
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
960 SPtyrValTyrValGluPro.....Ala 967
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1284 ACTGGAAACCTTCAGTCTGAGATGATACCTTTTGCAGGGGTTCAAAGAGA 1333
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
968 GluAlaAsnGluProHisIleValCysIleGluArgIleuTrpGln 984
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1334 TAAAGGAACAAAGTCACTCCAGCTGTATTATTGTGAAATTATGGGCGCTTACA 1383
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
984 PserAlaGlyIu.....LysTrpLeu...TyrGlyCysTrp 995
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1384 GTTCTTATGACCCGATATGATACCTCCACATTTTGGAAATATACACAAGGAT 1433
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
996 ..PheTyrArgProAsnGluThrPheHisLeuAlaThrArgLysPheLeu 1011
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1434 GATTTGTGATTAATCTATATCAACTTAAGTGGGAGAGACCTGATTTTCCAG 1483
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1012 GluArgGluValPheLysSerAspTyrTyrAsnLysValProValSerLys 1028
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1484 TGAATTCAGC.....ATCCATGAGATTTTGGCCAGCT 1515
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1028 sIleLeuGlyLysCysValValMetPheValLysGluTyrPheLysLeu 1045
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1516 GC.....CAAGATATATCCGATATGTCATGCGAGATTAATTACG 1553
      ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1045 YSProGluAsnPheArgAspGluAspValTyrValLysGluSerArgTyr 1061
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1554 GATGTTTAAACAAAGAGGGGCGCATTTCCAGGACCTCAAGACAGATGAGAT 1603
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1062 SerAlaLysThrLysSerPheLysLysIleLysLeuTrpThrMetProVal 1078
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1604 GTCA.....TTGCCT..... 1613
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1078 ILeuSerValArgPheValProArgAspValProLeuProValValArgVal 1095
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1614 .....GAAGATGAGGCCCTACT 1631
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1095 aAlaSerValPheIleAsnThrAspLysAlaGluGluGluLysSHSLeu 1111.
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1632 AGGACACTTGCACAGAGAAAGAAAGTACAGACAGATTACAGAAATAGAGCC 1681
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1112 AspThrLeuAspSerLysValGlyGluSerIleLeu..... 1124
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1682 ACCAGGGCGTTGCATCTCCAGTACTCAAGACAGGCTCATAGCGCTGAAG 1731
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1125 .....HisLeuGluLysAspLysGluAsp..... 1132
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1732 CAGTAACAAATTTGGCGCTTCCACTTGAAGTTTTCAGCTCGAA 1775
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1133 .....ValProValGluMetSerAsnGlyIu 1141
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

seq_name: pir2:T34036
hypochemical_block:
hypothetical protein B0041.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34036
R:Fulton, R.; Wohlmann, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid B0041.
A:Accession: T34036

```

A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-1359 <EUL>
A: Cross-references: EMBL:AF000196; PIDN: AAC24256.1; GSPDB: GN00019; CESP: B0041.7
A: Experimental source: strain Bristol N2; clone B0041
C: Genetics:
A: Gene: CESP: B0041.7
A: Map position: 1
A: Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

alignment_scores:
Quality: 206.50 Length: 660
Ratio: 0.599 Gaps: 23
Percent Similarity: 52.273 Percent Identity: 21.212

alignment_block:
US-09-687-230-1 x T34036 ..

Align seg 1/1 to: T34036 from: 1 to: 1359

```
129 GCGCGGGGGGGGGGCGACCGGGCCGGTGGCATGGCAGACAGACAA 178
      |||:::  ::  ::  ::|||:::  ::  ::||
86 AAllysSerGIuSerGIuSerGIuSerGIuSerGIuSerGIuSerGIu 102
      |||:::  ::  ::  ::|||:::  ::  ::||
179 GAAGCACAAGTCGAGACAAACCTCTACAGAGGATGTAGAAACCCCT 228
      |||:::  ::  ::  ::|||:::  ::  ::||
102 slysserlysserlysslyslvalspdlunlysllyslglulysserl 119
      |||:::  ::  ::  ::|||:::  ::  ::||
229 TGAAGCTGTCTCAAAATAGAGAGGAGACCGAAGTCACCACTCCACG 278
      |||:::  ::  ::  ::|||:::  ::  ::||
119 ysllys.....LysArgThrThr 124
      |||:::  ::  ::  ::|||:::  ::  ::||
279 GCGAGCTCGGGGCGACGACTCCAGCCCTTCGAAGACAAAAGCATCATGA 328
      |||:::  ::  ::  ::|||:::  ::  ::||
125 SerSerSer.....GluAspGluAspSerAsp 134
      |||:::  ::  ::  ::|||:::  ::  ::||
329 CAATCAACAGACAGACAAACGGAAGAAAGAAAGAGAGAGAAAGCAGA 378
      |||:::  ::  ::  ::|||:::  ::  ::||
134 uGIuAArgGIuInlysserlysslysserlysslystrlysllyslglnt 151
      |||:::  ::  ::  ::|||:::  ::  ::||
379 TTCACGGGGA.....GAAAGGGGAGAAAACGGAAGAGCTTAAGAG 422
      |||:::  ::  ::  ::|||:::  ::  ::||
151 hrSerSerGIuSerGIuSerGIuSerGIuSerGIuSerGIuSerGIu 167
      |||:::  ::  ::  ::|||:::  ::  ::||
423 GATTAATAAGACGAGATCGAGAC.....CGCGTGGAGAAATGAGGC 463
      |||:::  ::  ::  ::|||:::  ::  ::||
168 SerLyssLyssAmlyssGIuLysserValLysslyslvalglulhrSer 184
      |||:::  ::  ::  ::|||:::  ::  ::||
464 AGAAAAAGATTCACAGTCTACGCGCCCTGTGAGATTAGACTGCTCGTG 513
      |||:::  ::  ::  ::|||:::  ::  ::||
184 uGIuSerAsp.....GluAspGIuLyssPros 193
      |||:::  ::  ::  ::|||:::  ::  ::||
514 AGAAGCTCTTCACAGCTCTTAGCCAAACAGAGAGAGATAGAACAGACA 563
      |||:::  ::  ::  ::|||:::  ::  ::||
193 erLyssLysserLysslyslleuLyssLyssAlalysserGIuSerGIu 209
      |||:::  ::  ::  ::|||:::  ::  ::||
564 CCCCTTCAGAGACTTGAATCAACGATGAGAGACAAATTCACAGAAAAGA 613
      |||:::  ::  ::  ::|||:::  ::  ::||
210 SerGIuSerGIuAspGIuLyssGIu..ValLyssLysserLysslyssse 225
      |||:::  ::  ::  ::|||:::  ::  ::||
614 TCCAACTGCTTCTCTTCAATTCCTGTGACGATTTATTTGCTCCTGGCT 663
      |||:::  ::  ::  ::|||:::  ::  ::||
225 rlyslvalvallysllyslGIuSerGIuSerGIuSerGIuAspGluAlu 242
      |||:::  ::  ::  ::|||:::  ::  ::||
664 ACTCCATGATCATTAACACCCAAAGGATTTTAGTACCTGAAAGAAAG 713
      |||:::  ::  ::  ::|||:::  ::  ::||
242 ysllysllystrGIuLyssArgLyssArgSerLystrhrSerSerGIuSer 258
      |||:::  ::  ::  ::|||:::  ::  ::||
714 ATCAAGAAACAAATGACATGATTCATAGCAAGAAACAAATGATTAAC 763
      |||:::  ::  ::  ::|||:::  ::  ::||
259 SerGIuSerGIuLysserAspGIuGIuSerGIuGIuGIuLyssGIuSer 275
```

```

764 ACTAATGTGTACTAATGCCATGATTACATAAACAAGACCATTTATT 813
      |||||
275 O.....LysProLysLysLysLysP 282
      |||||
814 ATAAAGCTGCAAGAAAGCTGTG.....CACTACAGCA 845
      |||||
282 roLeuAlaValLysLysLeuSerSerAspLysLysSerGluLysSerAsp 298
      |||||
846 ATGAAATTTCTAGCCAGAAAGAAATTCAGAGCTGAGCAGACAGATAGA 895
      |||||
299 ValGluValLeuProGluLysLysLysArgGlyAlaValThrLeuIleSe 315
      |||||
896 CTTCATGGCTGACTGCAGAAAACTCGAAAGCAAGAAATGAAACAGACA 945
      |||||
315 rAspSerGluAspGluLysAspGluLysSerGluLysSerGluLysSer 332
      |||||
946 CCTCACAGAGTGGGAGAGACGAGCGCTGCGCAGAGAGAGAGAGAGAC 995
      |||||
332 ALGluGluLysValSerLysLysLysAlaLysLysGluLysSerGlu 348
      |||||
996 TCTGAGATGCCGAGACACAGCGCTTCAGAGATCCAGCAAGAAATAA 1045
      |||||
349 SerGlySerAspSerSerGluGlySerIleThrValAsnArgLysSer 365
      |||||
1046 AAGAAAGACAAACATATGCTTGAAGATAAG.....TTTAAAGCA 1086
      |||||
365 sLysLysGluLysProGluLysLysLysGlyIleIleMetAspSer 382
      |||||
1087 ATATATTAGAGAGAGACAGAGAGCTTGACCGCATCGTAAGAAATCT 1136
      |||||
382 eLysLeuGluLysGluThrIleAspAlaGluAlaGluLysGluArg 398
      |||||
1137 GGAGGAAGCTGACACGCGCTTGAGAACGTGACGTGCGAATTTGAAAG 1186
      |||||
399 .....ArgLysArgLeuGluLysGluLysGluPhe..... 409
      |||||
1187 AAGAAACACAGATGAAACACGACGTTGGACTTCTCCATCCTTGATC 1236
      |||||
410 .....AsnG 411
      |||||
1237 CCATTGTAGAGAGACGACGCTACTGCTGTGAGACTGAGATGACAAT 1286
      |||||
411 LysIleValLeuGluGluLysGluLysPheThrGlu.....MetLeuThr 425
      |||||
1287 GGA.....AGACTTCAGTCTGAGAGTGAATTAATCTTGGCAGGG 1321
      |||||
426 GlyThrSerSerGluArgLysLeuLysSerValValLeuAspProAspSe 442
      |||||
1322 GTTCAAGAGATAAAGAAAGCAAGTCACCTCCAGCTTATATTTGAAAT 1370
      |||||
442 rSerThrValAspGluLysSerLys..LysProValGluValHisAsnS 458
      |||||
1371 .....TATGGCCCTACAGTCTTATGACACCGCAT..... 1400
      |||||
458 eLysValArgIleLeuLysProHisGluAlaHisGlyIleGluPheMet 474
      |||||
1401 TATAGTCCACATTTGCAATATACAGAGATGATTCTGATTTAATGTA 1450
      |||||
475 TyrAspCysAlaPheGluSerLeuAspArgLeuAspThrGluLysSerG 491
      |||||
1451 TTCAACSTATGGGGAAGACSTGATCTTCCAAGTATTCAGATTCAGATG 1500
      |||||
491 yGlyIleLeuAlaHisCysMetGlyLeuGlyLysThrLeuGluValIleT 508
      |||||
1501 AGTTTGTGGSCASTGSCCAAGATTAATCCGTAATGTC.....ATGSCAGAT 1544
      |||||
508 hrPheLeuHisThrValLeuMetHisGluLysIleGlyLysCysLys 524
      |||||
1545 AGTTTACTGGATGTTTAAACAAAGAGAGGCAATTCAGAGACCTTACAGA 1594
      |||||
525 ArgValLeuValValArgLysAsnValIleIleAsnTrpPheLysGlu 541
      |||||
1595 GATGAGATGTGATTTGCGTGAAGATGAAAGCCATCTAGACACATTTGACA 1644

```

```

      |||||
541 uphGluLysTrpLeuValAspAsnAspGluGluLeuAspThrIleAsp 558
      |||||
1645 CAGGAAAGAAATGAGACAGATTAACAGAAATAGAGCAGCAGCGCGTTGG 1694
      |||||
558 AlaAsnGlu.....Leu 561
      |||||
1695 GACTCC.....AGTACTCAGACAGCGCTCATAGCGCTGAAACCA..... 1733
      |||||
562 AspSerLysThrIleGluAspArgArgAlaLeuLysAlaTrpH 578
      |||||
1734 .....GTAAACAAATTTGGCGTTCAGATGAAAGTTT 1764
      |||||
578 sSerSerLysThrProSerValMetIleIleGlyTrpAspLeuPheArg 595
      |||||
595 leuThrValGluAspAspProLysLysLysLysProLysAsnArgAsn 611
      |||||
1815 AGATTGCTCAGGCAACTCCAGGAAAGCCAGAAATGAAACGTTTGAGACAC 1864
      |||||
612 ArgArgLeuGluLysAlaLysGluAspPheArgLysThrLeuGluAsnPr 628
      |||||
1865 ACCGCCGGAACATGATCTGT.....CTCTGGGTCCT 1899
      |||||
628 OGlyProAspMetValValCysAspGluAlaHisLysLeuLysAsnAsp 645
      |||||
1900 CATCA.....GAGAAATGCAATC.....TTGCTGAACAGTGAACCAATA 1937
      |||||
645 sPseAlaLeuSerLysCysMetValLysIleLeuThrLysArgArgIle 661
      |||||
1938 ATCTTAAGAATTCACACAGCAATTAATCTCC 1967
      |||||
662 CysLeuThrArgLysThrProLeuGluAsnAsn 671
      |||||
seq_name: p1r2:S45251
seq_documentation_block:
SNF2alpha protein - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C:Accession: S45251
R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A:Title: Two human homologues of Saccharomyces cerevisiae SNF2/SNF2 and Drosophila br
A:Reference number: S45251; MUID:94268902
A:Status: preliminary
A:Accession: S45251
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1572 <CH>
A:Cross-references: GB:ID26155; NID:9505086; PIDN:BA05142.1; PID:dt1005684; PID:98766
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1409-1464/Domain: bromodomain homology <BRO>
alignment_scores:
Quality: 205.00 Length: 349
Ratio: 1.265 Gaps: 14
Percent Similarity: 46.418 Percent Identity: 23.496
alignment_block:
US-09-687-230-1 x S45251 ..
Align seg 1/1 to: S45251 from: 1 to: 1572
321 GATCATGACAAACACAGAGAC.....AGAAAGCGGAAAGAGAG 358
      |||||
1257 AspMetAspArgArgArgGluAspAlaArgAsnProLysArgLysProArg 1273
      |||||
359 AAGAAAGAGAGAGAGAGATTCAGAGG..... 386
      |||||
1273 GluMetGluGluAspGluLeuProSerTrpIleIleLysAspAspAlaG 1290

```

```

387 .....CAAGAAAG..... 395
1290 luValgluarleuthrhcysgluIngluIngluysllephglYarg 1306
396 ...GGGAGAAAACGAGA..... 410
1307 GlserrarglnrargaspValaspTyrSeraPalaleuthrGluIu 1323
410 ..... 410
1323 sgIntrPleuargAlalegluaspLysleuIngluIngluMetgluInglu 1340
411 .... AGAGTTAGGAGGATTAAGAAAGAGA.....GATCAGACCGG 449
1340 luValArtrleuLysLysArgLysArgArgArgValaspLysaspPro 1356
450 GTGGAGATGAGCAGAAAAGATCTCAGGTGCAGCCCGCTGTGAGAT 499
1357 AlalysgluaspValgluLys.....AlalysLysArgArg 1368
500 AGACTTGCTCTGAGAGAGCTCTCAGAGCTTTAGCCAAACAGAGAG 549
1368 gglYargProProAlaglLysleuSerProasnProProlys..... 1382
550 AAGTAGACAGACACCCCTTCAAGAGCTTTGATCAACAGTGATG...AGA 596
1383 .....LeuthrLysgluMetasnAlallelleaspThr 1393
597 CAATTCAGAGAAAAGATCCAGT.....GCTTCTT 628
1394 ValIleasnTyrLysaspSerSerGlyArggluInleuSerGluValrheII 1410
629 TTCATTCTGCTGAGCTGATTTATTCCTGCTGCTCTCAGTATCATTA 678
1410 eglInleuProSerArgLysgluInleu...ProgluTyrTyrgluInleuIIa 1426
679 AACACCCAGATGATTTTACTACATGAAGAAAAGATCAAGAACATGAC 728
1426 rglYerProValaspheLysLysLysLysLysLysLysLysLysLys 1442
729 TATCAGTCCATGAAAGAACTAAAGATTAATCTCAATGATGATTA 778
1443 TyrArgSerleuGlyaspLeuIngluLysaspValMetleuLeuLys 1459
779 TGCATGATTACATTAACAGACGACCTTTATATATAAGCTGCAGAGA 828
1459 nAlaeglIntrPheasnleuGlySerGlnleTyrGluaspSerlleV 1476
829 AGCTTGTCACACAGAAATGAAATTTTACGAGAAAGAAATTCAGAGC 878
1476 alleu..... 1477
879 CTGAAGCAGAGATGACTTCTGCTGCTGCTGCAAAAAGCTGCAGAGA 928
1478 .....GlnSerVal...PheLysSerAlaArgGlnLyslleAlaLysgl 1491
929 GAAAGATGAGACAGACACCTCAGAGATGGGAGAGAGAGAGGCTGCTGC 978
1491 ugluInleuSerGluaspGlnSerAsnGluInglu.....G 1503
979 AGAGAGAGAGAGAGAGCTGAGAGATGCCAGAGACAGAGCTTCAAGAGT 1028
1503 luIngluInleuSerGlnleuSerGlnleuSerGlnleuSerValLysVal 1519
1029 CCCAGCAAGAAAATTAAGAAAGAAAGCAAAATATGCTTGAAGATTAAT 1078
1520 LyslleLysleuasnLysLysaspLysglYargaspLysglYlysG 1536
1079 TAAAGC.....AATATTTAG 1095
1536 yLysLysArgProasnArgglYLysAlaLysProValalSeraspPheA 1553
1096 AGAGAGAGAGAGAGAGCTTACCGCATCTGAAGAAATTCGAGAGA 1142

```

```

seq_name: p1r2.T33328
1553 spseraspGluIngluInaspGluArggluInleuSerGluIngluysSerGly 1568
seq_documentation_block:
  hypothetical protein F13C5.2 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
  C:Accession: T33328
  R:Wohlmann, P.; Hawkins, J.; Gilliam, B.
  Submitted to the EMBL Data Library, July 1998
  A:Description: The sequence of C. elegans cosmid F13C5.
  A:Reference number: 221324
  A:Accession: T33328
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-374 <WH>
  A:Cross-references: EMBL:AF077531; PIDN:AA64610.1; GSPDB:GN00028; CESP:F13C5.2
  A:Experimental source: strain Bristol NZ; clone F13C5
  C:Genetics:
  A:Gene: CESP:F13C5.2
  A:Map position: X
  A:Introns: 25/3; 135/3; 189/3; 313/1
  C:Superfamily: bromodomain homology
  F:141-198/Domain: bromodomain homology <BRO>

alignment_scores:
  Quality: 204.00 Length: 365
  Ratio: 1.041 Gaps: 18
  Percent Similarity: 53.699 Percent Identity: 26.849

alignment_block:
  US-09-687-230-1 x T33328 ..

Align seg 1/1 to: T33328 from: 1 to: 374

256 ACGAAGTCACCGAACTCTCCAGCGGAGCTCGGGCAGACATCCAGCTC 305
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
4 ThrArgSerAlaAsnHisProAlaProlLysArgAlaIleSerProAsp 20
306 TTAGC.....AAGCAAAAAGATCATGACAAACA 334
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 rAspaspGluIngluInaspPheLysSerLys..... 31
335 CAAGACAGAAAGCGGAAAAGAGAAAGAAAGAGAGAAAGAGATTCAG 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
32 ..ArgThrProSerGly...ArgGlyArgGlyArgGlyArgGlyArg 46
365 GGGAGAAAAGGGA.....GAAAACGG 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
47 GlyArgGlyArgGlyGlyAlaThrAlaAlaGlyAlaThrValGluGly 63
408 AGAAGATTAGGAGGATTAAGAAAGAGAGAGATTCAGACCGGTGAGAA 457
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
63 yelSerThrArg.GlyGlyGlyThrGlyArg.....GlyGlyArg 76
458 TGAGGCAAGAAAAGATCTCAGAGTGCAGCCCGCTGAGATTAGACTTGC 507
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
77 ...GlyArgProArgGly.SerLysAsnSerAspGlyGlyAlaGlyGly 92
508 CTCTGAGAGAGCTCTCAGAAAGCTCTTACGCAACAGAA..... 548
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
92 roProAlaLysArgGlyArgGlyLysLysAlaLysSerGluSerGluasp 108
549 GAGTAGAAGACAGACCCCTCAGAAAGCTTTGATCAACTGATGAGAGCA 598
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 GluAlaGluSerAspHisLeuHisaspGluLeuLysLysCysLeuSer11 125
599 ATTGACAGAAAAGATCCAGTCTTTC.....TTTCAATTCCTGTG. 641
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 eleuLysgluInleuLysSerThrHisAspSerPheThrPheProPheA 142

```



```

308 yHhS...SerpHeGluTyrAspLysGluThrMetIleLysArgLeuLys 323
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
357 AGAAGAAAGAGAGAGAGAGATTCACGAGGGAAGAAAGGAGAAAG 406
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
324 LeuGluGluSerAspLysMetIleGluLysGlyLysLysLysSer 340
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
407 GAGAAGAGTTAAGAGAGATAAAGAGAGAGAGATGAGACCGGAGAG 456
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
340 gSerAspLeuGluAlaIleThrAspLeuGlnAspArgGluAsnThrAsn 357
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
457 ATGAGCAGAGAAAAGATCTCCAGTGTACGCCCCCT 491
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
357 spGluProAspThrAsnGlnLysLeuProThrProGluLysSerThrPhe 373
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
492 .....GTGAGATTGACTGTGCTCC 511
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
374 SerAspThrGlyAsnLysArgProLysGlnSerAsnLeuAspLeuThrVa 390
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
512 TGAG.....AGCCTCTCACAGCTCTTTAG 537
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
390 LAsnLeuGlyIleGluAsnLeuSerLeuLysThrLeuLeuSerSerIleG 407
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
538 CCAAAACA..... 545
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
407 LngLlnLysSerGlnLeuGlyIleSerAspTyrGlnLeuLysThrLeu 423
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
546 .....GAAGAACT 553
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
424 IleMetAspValArgLysAsnArgSerLysTrpThrSerAspGlnArgI 440
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
554 AGACAGACACCCCTTCAAGAGCTTTGAATCACTGATGAGACAATTGC 603
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
440 eGlyGlnGlnGluLeuTyrGlnLacLysGlnValValLeuGlnLeuVa 457
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
604 AAGA...AAAGATCCAGTCTTCTTCACTTTCCTGTGACTGATT 650
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
457 rGAsnTyrThrGlnLysSerThrProPheLeuAsnLysValSerLysArg 473
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
651 ATTGCTCTGCTACTCCATGATCATTAACACCAATGATGATTTAGTAC 700
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
474 GlnAlaProAsnTyrHisGlnIleIleLysLysSerMetAspLeuAsn 490
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
701 CATGAAGAAGATCAGACAAATGACTATCATGCTCATAGAAGACTTA 750
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
490 rValLeuLysLysLeuLysSerPheGlnTyrAspSerLysGlnGluPhe 507
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
751 AGGATAACTCAACTAATGTGTACTAATGCATGTTTACATATAACCA 800
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
507 aLAspAspIleMetLeuIleTrrPlysAsnCyLeuThrTyrAsnSerAsp 523
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
801 GAGACATTTATTATAAGCT...GCAAGAAGCTGTTCACCTCAGGAA 847
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
524 ProSerHisPheLeuArgLysHisAlaIleAlaMetGlnLysLysSer 540
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
848 GAAATTTCTTACCGAAGAAATTCAGAGCTGAAGCAGAGACATAGACT 897
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
540 uGlnLeuIleArgMetIleProAsnIleThrIleArgAsnArgAlaAsp 557
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
898 TCATGGCTGACTTGCAGAAACTCGAAAGCAGAAAGAT 935
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
557 euGlnLysGlnIleGlnAspMetGlnLysAspLysAspTyrGlnLeuAsp 573
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
936 ...GGAACAGACCTTCACAGAGTGGGAGAGACGA.....GGCTG 973
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
574 GlnGlnGlnGlnValAlaGlySerGlyArgLysGlyLeuAsnMetGlyAl 590
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
974 CTGGCAGAGAGAGAGAGAGACTCTGGAGATCCGAGACACAGCCTTCA 1023
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
590 aHisMetLeuAlaLysGlnLysGlnValSerGlnLysAspSerSerL 607
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1024 AGAGTCCCGCAAGAA..... 1040
      :::: :::: :::: :::: :::: :::: :::: :::: ::::

```

```

607 yStrValLysAspGlnAlaProThrAsnAspAspLysLeuThrSerVal 623
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1041 .....AATAAAAGAAAGCAAA..... 1058
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
624 IleProGlnGlyGlnLysGlnLysAspLysThrAlaSerSerThrValTh 640
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1059 .....GATATGCTTGAAGATTAAGTTTAAAGCAATAATTNAGA 1098
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
640 rValHisGlnAsnValAsnLysAsnGlnIleLysGlnAsnGlyLysAsnG 657
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1099 GAGACAGAGACAGCCTTGACCCGATCGTAAGACATCTCG.....AGC 1141
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
657 LngLlnLysPheValGlnGlnSerSerLysThrGlnAspSerSerLys 673
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1142 AAAGCTGACAGCGCGCTGTGACAGTCAGTCGCAATTTGAAGAAGA 1191
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
674 AspAlaAspAlaAlaLysLysAspThrGlnAspGlyLeu...GlnAsp 689
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1192 AACAGATGGAACAACAGAGCTGGAGCTTCTCATCTGTGATCCAT 1240
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
689 sThrAlaGlnAsnLysGlnLacLysGlnAsnGlnGlnGlnLysAspA 706
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1241 .....TGAGAGAGCCAGGCTA 1258
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
706 sPAspAspGlnAspGlnAspGlnAspMetValAspSerGlnSerTyrLeu 722
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1259 CTGCCTGTGAGACTGGGATGACAACTGAGACCTTCAGTCTGG..... 1303
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
723 LeuGlnLysAspAspAspArgAspAspLeuGlnIleSerValTrrPlysTh 739
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1304 .....AGTGAATACT 1313
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
739 rValThrAlaLysValArgAlaGlnIleCyLeuLysArgThrGlnTyrP 756
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1314 TT.....GCAGGGGTCAAGAGGA 1333
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
756 HeLysAsnGlnLysLeuAsnSerAspSerGlnAlaPheLeu...LysAsn 772
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1334 TAAAGCAACAAGTCACTCCAGTGTATATTGAAATATGGCCCTTCA 1383
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
772 oGlnArgMetLysArgPheAspGlnLeuPheLeuGlnLysGlnGlnL 789
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1384 GTTCTTATGACACCGATTTATGACTCCACATTTGCCAAT 1421
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
789 yAlaLeuGlnLeuSerTyrArgGlnLysIleGlnGlnAsnSerIleMetLys 805
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1422 .....ATCAGCAAGATGATTCGATTTAATC..... 1448
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
806 AsnGlyPheGlyThrValLeuLysGlnGlnLysAspAspGlnLeuGlnP 822
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1449 .....TATTCACCTATGGGGAAGACTCT.....GATCTTCCA 1482
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
822 eHisAsnAspHisSerLeuAsnGlnLysAlaIlePheGlnLysGlnPro 839
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1483 GTGATTTGAGCATCCATGAG.....TTTGGCCACGCGCAATATTA 1526
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
839 snAspIleGlnLeuAspAspThrArgPheLeu.....GlnIleTyr 852
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1527 CCGTATGTGATGAGATAGTTTACTGATCTTTAAACAAAAGAGGCA 1576
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
853 Asp.....IleSerAsnAlaIleProAspIleValTyrGlnGlyValAs 867
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1577 TTCAGAGACCTTCAAGAGATGAGATGTCAATGGCTGAAGTGAAGGCC 1626
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
867 nThrLysThrLeuAspLysMet.....GlnAspAlaServ 879
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1627 ATACTAGGACACTTGACACAGGAAAGAAATGAGACAGATTACAAAGTA 1676
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
879 aLAspArgMetLeuGlnAsnGly..... 886
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
1677 GAGCCACAGAGGCGTTTGACTCCAGTACTCAAGACAGGCTATGAGCT 1726
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
887 .....IleAsnLysGlnSerAlaArgPheLeuAlaAs 896
      :::: :::: :::: :::: :::: :::: :::: :::: ::::

```


THIS PAGE BLANK (USPTO)


```

1517 easnaspYrlyVsaspIle..... 1533
1690 GTTTGACTCCGACTACTCAAGAAGGTCTATAGCGCTGMAAGCACTAACA 17339
      ::::: ::::: ||| :::
1524 ..PhelysInlaIaaEngluAspArgLeuThrSerAlaIysGluLeuPro 15399
      ::::: ||| :::
1740 AATTTTGGCGTTCCAGTTGCAAGTTTGTGACTGCAAGAGCTGAAATAT 17899
      ||| ||| :::
1540 TyrPhe..... GluIysAspPheTr 1546
      :::::
1790 CCAGAAAGAACTTGATGAGACCAACAGATTGCTCAGGAACTCCAGAGAG 18339
      ::::: ||| :::
1546 pProasnValLeuGluGluSer.....IleYsGluLeuGluGluG 15606
      :::::
1840 CCCAGATGAGAGCT.....TGAGCACCAAGACC 18666
      ::::: |||
1560 LngluIngluGluGluYsLysGluGluSerThrAlaIaSerGluThrPro 15766
      :::::
1869 CCTGGAGACATGATCTCTCTCTGGGCTCCCTCAACAGAAATGCATCTT 1918
      ||| |||
1577 GluGlySerGln.....GlyAspSer..... 15833
      :::::
1919 GCTGAACACAGTGCACATATCTTAAGATTTGCACAGCAAGTACCTCCA 19666
      ||| ||| :::
1584 .....LysAsnAlaLysIysLysAsnL 1592
      :::::
1969 GGATATCTGTAGCACGTATGAGT...TCGAAAGCAATGGGATTTTC 2015
      ::::: ||| ||| :::
1592 yAlsthrsInLysAsnLysSerSerIleSerArgAlaAsnLysLysLys 1608
      ||| :::
2016 CATTCCTTCCCGCTCATGGAAAAACAACCTT 2045
      ||| :::
1609 ProSerMetProAsnValSerAsnAspLeu 1618
      :::::
seq_name: /cgn2_6/plodata/2/1aa/6A.COMB.pep:US-08-961-739-2

seq_documentation_block:
? Sequence 2, Application US/08961739A
? Patent No. 6063583
? GENERAL INFORMATION:
? APPLICANT: Montelim, Marc R.
? TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
? FILE REFERENCE: SAK1650-1
? CURRENT APPLICATION NUMBER: US/08/961,739A
? CURRENT FILING DATE: 1997-10-31
? EARLIER APPLICATION NUMBER: US 194,468
? EARLIER FILING DATE: 1994-02-10
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 2
? LENGTH: 2441
? TYPE: PRT
? ORGANISM: Mus
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: (1)...(2441)
? OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

alignment_scores:
      Quality: 227.50      Length: 894
      Ratio: 0.625      Gaps: 42
Percent Similarity: 40.716      Percent Identity: 20.470

alignment_block:
US-09-687-230-1 x US-08-961-739-2 ..
Align seg 1/1 to: US-08-961-739-2 from: 1 to: 2441

11 ATGCGCCCGCGCGCGCGCGCGCG.....CCCGCTGCTGCGCGCGCGG 54
:::||||| ::::: ||| ||| ||| ||| |||
838 LeuAlaProGlnIaaSerGlnLeuProCysProValThrGlnSerPr 854

```



```

1392 yfargThrIysAlaLeuPheAlaPheGluIleAspGlyValAspVal 1408
1410 ACATTTCGAATATATACGACGACGATTCGATTATC..... 1448
1409 CysPhePheGlyMetHisValGlnAspThrAlaLeuLeuAlaProHisG1 1425
1449 .....TATTCACCTATGGGGAAGACTGTGATCTTCCAA 1482
1425 nileGlnGlyCysValTyrIleSerTyrLeuAsp..... 1436
1483 GTGATTTCAGCATTCATGAGTTTGGCCACGTGC..... 1517
1437 .....SerIleHisPhePheArgProArgCysLeuArgThrAlaVal 1450
1518 .....CAAGATTATCCGTA 1531
1451 TyrHisGluIleLeuIleGlyTyrLeuGluTyrValLysLysLeuValTy 1467
1532 TGTGATGGCAGATAGTTTACTGGATGTTTACAAAGAGGG..... 1574
1467 rValThrAlaHisIleThrPalaCysProProSerGluGlyAspAspTyrI 1484
1575 .....CATTCAGAGCCCTA 1589
1484 IepHeHisCysHisProProAspGlnLysIleProLysProLysArgLeu 1500
1590 CAAGAGATGAGATGTCATTGCTCCTGAGATGAGAGCCATCTAGACACAT 1639
1501 GlnGluTyrTyrLysLysMetLeuAspLysAlaPheAlaGlnArgIleI 1517
1640 TGACACAGAGAAAGAAATGAGCAGATTCACAGATGAGCCACGAGGC 1689
1517 eAsnAspTyrLysAspIle..... 1523
1690 GTTGGACTCCAGTACTCAAGACGGCTCATAGGCTGAAGAGCAGTACA 1739
1524 ..PheLysGlnAlaAsnGlnAspArgLeuThrSerAlaLysGluLeuPro 1539
1740 AATTTTGGCGTTCCAGTGAAGTTTTCGACTGTGAAGAGCTGAATATAT 1789
1540 TyrPhe.....GluGlyAspPheTr 1546
1790 CGAGAGAAACTTGATGAGACCCAGATTCGTCAGGAGAACTCCAGAGAG 1839
1546 ProAsnValLeuGlnLysSer.....IleLysGluLeuGlnG 1560
1840 CCCAGATGAACT.....TTGAGCACAGAGACC 1868
1560 IuGlnGlnLysLysLysGlnLysSerThrAlaAlaSerGluThrPro 1576
1869 CCTGGACATGATCTGTCTTGGTCCCTCATGACAGAGAAATGCAATCT 1918
1577 GluGlySerGln.....GlyAspSer..... 1583
1919 GCTGAACAAGTACCAATATCTTAAAGATTCGACAGCAATACTCA 1968
1584 .....LysAsnAlaLysLysLysAsnAsnL 1592
1969 GGTGATATCGTAAGCAGTATGAGT...TCGAAAAGCAATGGGATTC 2015
1592 yslYThrAsnLysAsnLysSerIleSerArgAlaAsnLysLys 1608
2016 CATTCCTTCCCGCTGACGAGAAACACTT 2045
1609 ProSerMetProAsnValSerAsnAspLeu 1618
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-227-536-2
seq_documentation_block:
; Sequence 2, Application US/08227536
; Patent No. 5658784
; GENERAL INFORMATION:

```

```

; APPLICANT: Eckner, Richard
; APPLICANT: Ewen, Mark
; APPLICANT: Livingston, David
; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227,536
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: DFCI-308XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-227-536-2

alignment_scores:
    Quality: 216.00      Length: 740
    Ratio: 0.663         Gaps: 33
    Percent Similarity: 44.054    Percent Identity: 21.216

alignment_block:
US-09-687-230-1 x US-08-227-536-2 ..
Align seg 1/1 to: US-08-227-536-2 from: 1 to: 2414

17 CGCGCCGGC.....CGCGCGCCCGCCCTGCTCGCGCGCGG 54
|||||
870 ProProGlnProGlnSerGlnAlaLeuHisProProPro..... 882
55 GGTCTGGGGGGCGGCTCCCGCCCTCGCGCTGCGCGCGGAGCGGGA 104
|||||
883 ....ArgGlnThrProThrProThrThrGlnLeuProGlnGlnValG 898
105 GCGGCGCCCGCAGCGCGCTGGCGCGCGGGGCGGCGGAGCGGCGCCG 154
|||||
898 InProSerLeuProAlaAlaProSerAlaAspGlnProGlnGlnGlnPro 914
155 GTGCGACATGGCGAGACACAGAGCAAGCAAGTC..... 190
|||||
915 .....ArgSerGlnGlnSerThrAlaAlaSerValProThrProAsnAl 929
191 .....GACAAACACCTCTACGA.....GGAGTAT 215
929 aProLeuLeuProProGlnProAlaThrProLeuSerGlnProAlaValS 946
216 GTAGAGAAAGCCCTTGAAGCTGCTCAAGATGAGAGAGGAGCAAGTAC 265
|||||
946 erIleGlnGlnGlnValSerAsnProProSer.....ThrSerSer 959
266 CGAACTCTCCACGGGCGAGCTCGGGGCGAGACTCCAGCCTCTTGAGAGCA 315

```

```

960 ThrGluValAsnSerGlnAlaIleAlaGluLysGlnPro..... 972
316 AAAAGATCATGACAAACAGAGACAGAAAGCGAAGAAAGAAAGAAA 365
973 .....SerGlnGluValLysMetGluAlaLysMetGluVal 984
366 GGAGAGAGACGATTCACAGGGGAGAGAGAGAGAGAGAGAGAGAGAG 415
984 AlaSrgLInProGlnProAlaAspThrGlnProGlnLysPrlAspL 1000
416 TAAGAGGATATAAAGAGCGAGATCGAGACCGGTCGAGATAGCGAG 465
1001 LysValGlnAspCysLysMetGlnSerThrGluThrGlu...GluArgS 1016
466 AAAAGATCTCCAGTGTACAGCCCTGTGAGATTAGACTGCTGCTGAG 515
1016 erThrGluLeuLysThrGluIleLysGluGluLysPrlAspLInPro 1032
516 AAGCTCTCACAGCTCTTTAGCC.....AAACAGAGAGAGAGT 553
1033 SerAlaThrGlnSerSerProAlaProGlnSerLysLysLysIlePh 1049
554 AGAAGACAGACCCCTTCAAGAGCTTTGATCACTGATGAGACAAATTGC 603
1049 elysProGluGluLeuArgGlnAlaLeuMetCProThrLeuGluAla 1066
604 AGAGAAAGATCCAGATGCTTTC...TTTTCATTTCTGTGACTGATT 650
1066 yArgGlnAspProGlnSerLeuProPheArgGlnProValAspProIn 1082
651 ATT.....GCTCTGCTACTCCATGATCATTAACACCCAGATGATT 694
1083 LeuLeuGlyIleProAspLysPheAspIleValLysSerProMetAsp 1099
695 TAGTACATGAAGAAAGATCAAGACATGACTAGTACGATCCATAGAG 744
1099 uSerThrIleLysArgLysLeuAspThrGlyGlnIleGlnGluProTrp 1116
745 AACTAAGATTAATCTCAACTAATGTACTAATGCCATGATTTACAT 794
1116 InTyValAspAspIleThrLeuMetPheAsnAsnAlaIleThrLeu 1132
795 AAACGAGAGACATTTATATTAAGCTCAAGAGAGCTGTGCTGACAG 844
1133 ArgLysThrSerArgValIleLysLysIleLysSerLysLeu..... 1145
845 AATGAAATTTCTTAGCCAGAGAAATTCAGAGCCTGAAGCAGAGCAT 894
1146 SerGluValIleGlnGlnGlu...IleAspProValMetGlnSerLeu 1161
895 ACTTCATGGCT..... 905
1161 LYTTCYSGYGLYARGLYSLEUGLUPHESERPROGLINTHRLAUCYSG 1177
906 ...GACTTGCAGAAAACTCGAAAGCAGAAATGAGACAGACACTCA 952
1178 TyrgLysGlnLeuCysThrIleProArgAspAlaThrIleTyrSer 1194
953 GAGTGGGAGAGACGAGAGCTGC.....TGGCAGAGAG 984
1194 rGlnAsnArgIleHisPheCysGluLysCysPheAsnGluIleGlnGly 1211
985 AGAGAGAGACTGTGAGATGCCGAGACAGACGCTTCAGAGATCCAGC 1034
1211 lSerValSerLeuGlnLysAspSerGlnProGlnThrThrIleAsn 1227
1035 AAAGAAAT...AAAAGAAAGACAAATATGCTGTAAGATTAAGTT.. 1079
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGluLeu 1244
1080 ...AAAAGCAATATTTAGAGAGACGAGACGAGCTT..... 1115
1244 eValGluCysThrGluCysGlyArgLysMetHisGlnIleCysValLeu 1261
1116 .....GACCCGATGCTGAG 1130
1261 lHisGluIleIleThrProAlaGlyPheValCysAspGlyCysLeuLys 1277
1131 GAATCTGCA..... 1139
1278 LysSerAlaArgThrArgLysGluAsnLysPheSerAlaLysArgLeu 1294
1140 .....GAAAGCTGACAGCGGCTTGTGACAGTACGTGCG 1176
1294 oSerThrArgLeuGlnIleThrPheLeuGluAsnArgValAsnAspPhe 1311
1177 AATTGAAAGAGAGAAACAGATGAGAACAGAGCTGGGCTTCAT 1226
1311 rGArgGlnAsnHisProGlnSerGlyGluValIleThrValArgValHis 1327
1227 CCTGTGATCCCATTTGAGAGAGCCAGGCTTACCTGCTGAGACTGCG 1276
1328 AlaSerAspLysThrValGlu.....ValLysProGln 1338
1277 AATGACAACTGAGAGACTTCAGTGTGAGATGATTTGACAGGGTTCA 1326
1338 yMetLysAlaArgPheValAspSerGly...GluMetAlaGlnSerPhe 1353
1327 AAGAGATTAAGAGACAAAGTCACTCCAGTGTATATTGAATTAGGG 1376
1353 ..... 1353
1377 CCTTACAGTTCTTATGCACCGCATTTAGTCAATTTGCAATATACG 1426
1354 ProTyrArgThrLysAla.....LeuPheAlaPheGlnGlu 1365
1427 CAAGATGATCTGATTAATCAT.....TCAACCTATG 1461
1365 uLleAspLeuValAlaLysCysPhePheGlyMetHisValGlnGluTyr 1392
1462 GGAAGACTGTGATCTTCAAGAT..... 1487
1382 LysSerAspCysProProProAsnGlnArgValIleLysSerTyrLeu 1398
1488 TTCAGCATCCATGAGTTTGGCCAGCTGC.....CAAGTTATCC 1528
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyr 1415
1529 GTATGTCATGCGACATGATTACTGATGTTTAAACAAAGAGGCGCAT 1578
1415 sGluIleLeuIle...GlyTyrLeuGluTyrValLysLysLeuGluTyr 1431
1579 CCAGGACCTTCAAGAGATGAGATGTCATTCCTGAAGATGAAGGCCAT 1628
1431 hr.....ThrGlyHis 1434
1629 ACTAGGACACTTGACACAGAGAAATGAGCAGATTACAGAAATGAGA 1678
1435 lLeThrAlaCysProProSerGlnGluLysAspAspTyrIlePheHis 1451
1679 GCCACAGGGCGT..... 1691
1451 sProProAspGlnLysIlePheProLysArgLeuGlnGluTyrTyr 1468
1692 .....TTGGAC..... 1697
1468 yLysMetLeuAspLysAlaValSerGluArgIleValHisAspTyrLys 1484
1698 .....TCCAGTACTCAAGACAGGCTCATACCGCTGAAGACAGT 1735
1485 AspIlePheLysGlnAlaIleThrGluAspArgLeuThrSerAlaLys 1501
1736 AACAAATTTGGCGTCCAGTGAAGTTTTCAGTCTGGAAGAGCTGAA 1785
1501 uProTyrPhe.....GluGlyAsp 1508

```



```

1178 TyrGlyLysGlnLeuCysThrIleProArgSpAlaThrTyrTyrSerTy 1194
953 GAGTGGGAGAGCGAGGCTGC.....TGCGAGAG 984
1194 rGlnAsnArgThrHisPheCysGlnLysCysPheAsnGlnIleGlnGly 1211
985 AGAAGAGAGACTCTGGATGCCGAGACACGCCCTTCAGAGTCGCCAGC 1034
1211 IuSerValSerLeuGlnLysAspProSerGlnProGlnThrThrIleAsn 1227
1035 AAGAAAT.....AAAAGAAAGACAAAGATATGCTTGAAGATAGTT.. 1079
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGlnLeuPh 1244
1080 ....AAAAGCAATATATTAGAGAGAGACGAGCAGCTT..... 1115
1244 eValGlnCysThrGlnCysGlnArgLysMetHisGlnIleCysValLeuH 1261
1116 .....GACCCGATCGTGAAG 1130
1261 IShISglnIleIleTrpProAlaGlyPheValCysAspGlyCysLeuLys 1277
1131 GAATCTGGA..... 1139
1278 LysSerAlaArgThrArgLysGlnAsnLysPheSerAlaLysArgLeuPr 1294
1140 .....GMAAGCTGACCGAGCGGCTTGTGAACGATGATGCG 1176
1294 oSerThrArgLeuGlnLysThrPheLeuGlnAsnArgValAsnAspPheLeuA 1311
1177 AATTGAAGAAAGAAACCAAGATGGAACAGACGTTGGGACTTCATCAT 1226
1311 rGArgGlnAsnHisProGlnSerGlnLysValThrValArgValAlaHis 1327
1227 CCTGTGATCCCATTTAGAGAGACGAGCTACTGCTGTGAGACTGGG 1276
1328 AlAserAspLysThrValGlu.....ValLysProGln 1338
1277 AATGACAACCTGGAAGACTTCACTGTGAGTGAATCTTGCAGGGGTTCA 1326
1338 YMetLysAlaArgPheValAspSerGly...GluMetAlaGlnSerPhe. 1353
1327 AAGAGATAAAGAACAAAGTCACTCCAGTTATATTGAATTATGGG 1376
1353 ..... 1353
1377 CCTACAGTCTTATGCACCGCATATATGATCCATTTGCAAAATATCAG 1426
1354 ProTyrArgThrLysAla.....LeuPheAlaPheGlnGln 1365
1427 CAGGATGATTCGTGATTTAATCTAT.....TCAACCTATG 1461
1365 uIleAspLysAlaLysPheLysPheGlyMetHisValGlnGlnLysTyrG 1382
1462 GCGAAGACTCTGATCTTCCACAGTAT..... 1487
1382 LysSerAspCysProProProAsnGlnAlaGlyValTyrIleSerTyrLeu 1398
1488 TTCAGACATCATGAGTTTGGCCACGTCG.....CAAGATTATCC 1528
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyrHis 1415
1529 GTATGTCAATGGCAGATGATTACTGATGTTTAAACAAAGAGGGCAT 1578
1415 sGlnIleLeuLeu...GlyTyrLeuGlnTyrValLysLysLeuGlnTyrT 1431
1579 CCAGGACCCCTAACAGAGATGAGATGTCATGCCGAGAGATGACGGCAT 1628
1431 hr.....ThrGlnHis 1434
1629 ACTAGACATTTGACACAGAAAGAAATGAGACAGATTAACAAGATAGA 1678
1435 IleTrpAlaCysProProSerGlnGlnLysAspArgThrIlePheHisCysHis 1451

```

```

1679 GCCACACAGGGCT..... 1691
1451 sProProAspGlnLysIleProLysProLysArgLeuGlnGlnTyrL 1468
1692 .....TTGGAC..... 1697
1468 YLysMetLeuAspLysAlaValSerGlnArgIleValHisAspTyrLys 1484
1698 .....TCCAGTACTCAAGACAGGCTCATAGCGCTGAAGACGT 1735
1485 AspIlePheLysGlnAlaThrGlnAspArgLeuThrSerAlaLysGlnLe 1501
1736 AACAAATTTGGCGTCCAGTTGAAGTTTTCAGCTTGAGACAGCTGAA 1785
1501 uProTyrPhe.....GlnGlyAsp 1508
1786 TATTCACAGAAACACTGATGAGACACACAGATGCTCAGGAGACTCCAG 1835
1508 hettProAsnValLeuGlnGlnUser.....IleLysGlnLeuGln 1521
1836 GAAGCCCGAATGAACGT 1853
1522 GlnGlnGlnGlnGlnGlnArg 1527

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-188-582-14

seq_documentation_block:
; Sequence 14, Application US/08188582
; Patent No. 5534410
;
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comali, Lucio
; APPLICANT: Dynalact, Brian D.
; APPLICANT: Hoev, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Neoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

US-08-188-582-14

alignment_scores:

Quality: 181.50 Length: 561
 Ratio: 0.717 Gaps: 21
 Percent Similarity: 45.098 Percent Identity: 21.034

alignment_block:

US-09-687-230-1 x US-08-188-582-14 ..

Align seg 1/1 to: US-08-188-582-14 from: 1 to: 1872

```

177 AAGAGCCACAGTCGACAAACCTCTACGAGCATATGTAGAGACC 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1354 LysLysArgArgValGlyThrThrValHisCysAspTyrLeuAsnArgpr 1370
227 CTTGAGCGTGGTCCCAAGTAGAGAGGAAACGACACCGAATCTCCCA 276
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1370 ohLysSerLeuHisArgArgArgThrAspProMetValThrLeuSerS 1387
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
277 CGGGCAGCTCGGGGACGACCTCCAGCTCTCGAAGACAAAACGATCAT 326
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1387 erLeuLeuSerLeuHisArgArgArgThrAspProAsnThrTyr 1403
327 GACAAACACAGGACAGAAACGGGAAAAGAGAAAGAGAGAGACCA 376
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1404 ProPheHisThrProValAsnAlaLysValLysAspTyrTyrLysI 1420
377 GATTCAGGAGGAAAGGAGGAGGAAAGGAGAGAGATTAAAGAGATTA 426
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1420 elle.....ThrArgProMetAspLeuGlnThrLeuArgLysAsn 1434
427 AAAAGAGAGCA.....GATCGAGACCG..... 449
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1434 alArgLysArgLeuTyrProSerArgGluGluPheArgLysHisLeuGlu 1450
450 .....GTGACAAATGACGACGAAAAAGATCTCCAGTGTACAGCCCTGT 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1451 LeuLeuValLysAsnSerLeuThrTyrAsnGlyProLysHisSerLeuTh 1467
494 GAGA.....TTAGACTTGCCCTCTGAGAGACCTCTGCACAA 528
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 rGlnHisSerGlnSerMetLeuAspLeuGlyAspGluLys..... 1480
529 GCTCTTTAGCCAAACAGAAAGAGTAGAAACAGACCCCTCAAGAGAGT 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1481 .....LeuLysGluLysGluAspLysLeuAlaArgLeuGluLysAla 1494
579 TTGCATCACTGATGAGACAATTGCGAGACAAAAGATCCAAAGTCTTC.. 626
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1495 IleAsnProLeu.....LeuAspAspAspGluValAlaIlePheSe 1508
626 ..... 626
1508 rPheIleLeuAspAsnIleValThrGlnLysMetMetAlaValProAspS 1525
627 .....TTTCATTCTCTGACTGATTTATTGCTCTCTGCGTACTCC 668
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1525 erTyrProPheHisProValAsnLysLysPheValProAspTyrTyr 1541
669 ATGATCATTTAAACCCCATGATTTTACCATGAAAGAAAAGATGCA 718
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1542 LysValIleValAsnProMetAspLeuGluThrIleArgLysAsnIleSe 1558
719 GAACAATGATATGATTCATAGAGAACTAAAGACTTAATCTCAACTCA 768
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1558 rLysHisLysTyrGlnSerArgGluSerPheLeuAspAspValAsnLeuI 1575
769 TGTGTACTAATGATCATGATTACATTAACAGAGACATTTATATATAA 818
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1575 leLeuAlaHisSerValLysTyrAsnGlyProGluSerGlnTyrThrLys 1591

```

```

819 GCTGCAAAGAGCTGTTCACATCGAGAAATGAAAATTCCTAGCCAG..... 863
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 ThrAlaGlnGluIleValAsnValCysTyrGlnThrLeuThrGluTyrAs 1608
864 .GAAGAAATTCAGAGCCTGAG..... 884
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 pGluHisLeuThrGlnLeuGluLysAspIleCysThrAlaLysGluAla 1625
885 .....CAGAGCATGACTTCATGCTGACTGTCGAG 914
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1625 leLeuGluGluIleValGluLeuGluSerLeuAspProMetThrProGluPro 1641
915 AAATCTCGAAAGCAGAAAGATGGAACACACACCTCAGACAGT..... 956
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 TyrThrProGlnProProAspLeuTyrAspThrAsnThrSerLeuSerne 1658
957 .GGGAGGAGCAGGAGCCTGCTGAGAGAGAGAGAGAGACTCT.....G 999
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1658 tSerArgAspAlaSerValPheGlnAspGluSerAsnMetSerValLeuA 1675
1000 GAGATGCCGAGACACACGCTTCAGAGTCCCGACAAAGAAATAAAGAG 1049
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1675 spIleProSerAlaThrProGluLysGlnValThrGlnGluGluAsp 1691
1050 AAAGACAAAGAT..... 1061
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1692 GluAspGlyAspLeuAlaAspGluGluGluThrValGlnGlnProGlu 1708
1062 .....ATGCTTGAGATGATGTTAAAGCATTAATTAGAGAG 1101
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1708 nAlaSerValLeuTyrGlnAspLeuLeuMetSer.....GluGlyG 1722
1102 AGCAGAGGACGCTTGACCCGATCTGTAAGAGAACTTGAGAGAGAC 1151
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1722 LysAspGluGluAspAlaGlySerAspGluGluLysAsn..... 1736
1152 AGCGCGCTGTGACAGCTCAGTCCGAATTTGAAAGAGAAACCAATAG 1201
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 .....ProPheSerAlaIleGlnLeuSerGlnSerGlySer..... 1750
1202 AACCAACGAGCTTGAGACTTCCATCTGATCCGATTTGAGAGAGAC 1251
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1750 rAspValGlySerGlyLysLeuArgProLysGlnProArgMet..... 1764
1252 CAGGCTACTGCTGTGAGACTGGGAATGACAACTGGAAGACTTCAGTCT 1301
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1765 .....LeuGlnGlu 1767
1302 GGAATGATATCTTTCAGGGGTTCAAAGAGATAAAGAGACAAAGTCA 1351
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 .....AsnThr.....ArgMetAspMetGluAsnGluLysE 1778
1352 TCCAGTGTATAT..... 1364
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1778 rMetMetSerTyrGluGluLysAspGlyGlyLysAlaSerHisGlyLeuGluA 1795
1365 .....TTGAATTAAGGCCCTACAGTCTTATGACACCGCAT..... 1400
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1795 spSerAsnIleSerTyrGlySerTyrGluGluProAspProLysSerAsn 1811
1401 ...TATGACTCCACATTTGCCAATATC..... 1424
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1812 ThrGlnAspThrSerPheSerIleGlyTyrGluValSerGluGlu 1828
1424 ..... 1424
1828 uGluGluAspGluGluGluGluGluGluGluArgSerGlyProSerValLeuS 1845
1425 .....AGCAAGATGATTCGATTTAATCTAATTC 1454
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1845 erGluValHisLeuSerGluAspGluGluAspSerGluAspPheHisSer 1861
1455 ACCTATGGGGAACACTCTGATCTTCCAACTGAT 1487

```

1862 ilealagly...AapserAspleuaspserasp 1871

seq_name: /cgn2_6/prodata/2/1aa/5a_COMB.pep:US-08-646-715-14

seq_documentation_block:

Sequence 14, Application US/08646715
Patent No. 5637686

GENERAL INFORMATION:

APPLICANT: Tjian, Robert

APPLICANT: Comai, Lucio

APPLICANT: Dynlacht, Brian D.

APPLICANT: Hoey, Timothy

APPLICANT: Ruppert, Siegfried

APPLICANT: Tanese, Naoko

APPLICANT: Wang, Edith

APPLICANT: Weinzierl, Robert O.J.

TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HONBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/646,715

FILING DATE: 09-MAY-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/188,582

FILING DATE: 28-JAN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 1872 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-646-715-14

alignment_scores:

Quality: 181.50

Ratio: 0.717

Percent Similarity: 45.098

Percent Identity: 21.034

alignment_block:

US-09-687-230-1 x US-08-646-715-14 ..

Align seg 1/1 to: US-08-646-715-14 from: 1 to: 1872

177 AAGAGAGCAAGTCGAGCAAAACACCTCAGAGGATATAGAGAGCC 226

1354 LysLysArgValGlyThrValHisCysPryLeuAsnArgPyr 1370

227 CTGAAGCTGTCTCAAAAGTAGAGGAGAACGAACTACCGAACTCTCA 276

1370 OHLSYSSerLEHLSArgArgThrAspPrometValThrLeuSers 1387

277 CGGCGACCTGGGGCAGCACTCCAGCCTCTTCAGAGCAAAAGCATAT 326

1387 erLELeuGluSerLEleAsnAspMetArgAspLeuProAsnThrTyr 1403

327 GACAAACACAGACAGAAAGCGGAAAAAGAGAAAGAAAGAGAGAGCA 376

1404 PROPHenIsthrProValAsnAlaLysValValLysAspTyrTyrSIL 1420

377 GATTTCAGGGGAGAGAAAGGAGAGAAACGAGAGAGTTTAAAGAGATA 426

1420 eile.....ThrArgPrometAspLeuGlnThrLeuArgLysnV 1434

427 AAAAGAAAGCA.....GATCGAGACCG..... 449

1434 alArgLysArgLeuTyrProSerArgLysGlnPheArgLysLeuGlu 1450

450GTGAGATGAGCGGAGAAAGATCCAGGTACGCCCTGT 493

1451 LeuLEValLysAsnSerAlaThrTyrAsnGlyProLysHisSerLeuTh 1467

494 GAGA.....TTAGACTTGCCCTCTGAGAGCCCTTCACAA 528

1467 rGlnLLeSerGlnSerMetLeuAspLeuLysAspGlnLys..... 1480

529 CCGCTTTAGCCAAACAGAAAGATAGACAGACACCCCTTCAGAAAGCT 578

1481LeuLysGlnLysGlnAspLysLeuAlaArgLeuGlnLysAla 1494

579 TTGAATCACTGATGACAAATTCAGAGAAAGATCCAGTGGTTTC. 626

1495 IleAsnProLeu.....LeuAspAspAspGlnValAlaPheSe 1508

626 626

1508 rPheLELeuAspAsnLEValThrGlnLysMetMetAlaValProAspS 1525

627TTTTCAATTCCTGTGACGTGATTTATGTCTCTGCTGACTCC 668

1525 erTTrProPheHisnIspProValAsnLysPheValProAspTyrTyr 1541

669 ATGATCATTTAAACCCCAATGATTTAGTAAACCTGAAGAAAGATCA 718

1542 LysValLEValAsnPrometAspLeuGlnThrLEArgLysAsnLESe 1558

719 GAACATGATATCATGTCATAGAGAACTAAAGATTAACCTCAAACTAA 768

1558 rLysHisLysTyrGlnSerArgLysSerPheLeuAspValAsnLeuI 1575

769 TGCTGCTAATGCCATGATTTACATTAACCCAGAGACCATTTATTA 818

1575 LeuValAsnSerValLysTyrAsnGlyProGlnSerGlnTyrThLys 1591

819 GGTGCAAGAAAGCTGTGCACTCGAATGAATAATGTTGCCGAG..... 863

1592 ThrAlaGlnGlnLEValAsnValCysTyrGlnThrLeuThrGlnLys 1608

864 .GAAAGAAATTCAGACCTGAG..... 884

1608 rGlnHisLELeuThrGlnLeuGlnLysAspILECysThrAlaLysGlnAla 1625

885CAGAGCATTAAGCTTCATGCTGACTGCTGACAG 914

1625 lAleuGlnGlnAlaGlnLELeuGlnLysSerPrometLThrProGlnPro 1641

915 AAACCTGGAAGAGCAAGAAAGATGGAACAGACACCTCACAGAGT..... 956

1642 TyrThrProGlnProProAspLeuTyrAspThrAsnThrSerLeuSerMe 1658

957 .GGGAGAGCAGAGAGCTGTGCGCAGAGAGAGAGAGAGACTCT.....G 999

1658 tSerArgAspAlaSerValPheGlnAspGlnLysSerAsnMetSerValLeuA 1675

```

1000 GAGATGCCAGACACAGCCCTCAGAGTCCAGCAAGAAATATAAAG 1049
      ::::: ||::: ::::: ||:::
1675 spileproseralathrproglulysglnvalthrglnlucgluas 1691
1050 AAGACAAAGAT..... 1061
      ||| |||
1692 g1yaspgllyasp1euln1asp1lucglulglthrv1a1gln1p 1708
1062 .....ATGCTTGAGATTAAGTTTAAAGCAATATTAGAGAG 1101
      ::::: ||::: ||::: ||::: ||:::
1708 n1alaserVal1eurygluaspleu1euser.....GluGlyG 1722
1102 AACGAGAGACGCTTGACCGCATCGTGAAGAAATCTGAGAGAACTGACC 1151
      ||::: ||::: ||::: ||::: ||:::
1722 luaspap1lucgluaspa1ag1yserasp1lucglulgluaspsn 1736
1152 AGCGGCTGTGACACATGCGCAATTTGAAAGAAACCAAGATG 1201
      ||| ||| ::::: ||::: ||:::
1737 .....Propheser1a1leg1n1euserglusergluserasp 1750
1202 AACACGAGACGCTTGACCTTCATCCATCCGATCCATTGAGAGAGC 1251
      ::::: ||| ::::: ||| ::::: |||
1750 r1aspval1g1ysergl1yletrp1ro1ysglnp1ro1ar1gme 1764
1252 CAGGCTACTGCTGTGACATGCGAATGACAACTGGAAGACTTCAGTCT 1301
      ||::: ||::: ||::: ||::: ||:::
1765 ..... 1767
1302 GGAGTCAATACCTTGACGAGGTTCAAGAGATTAAGAAACAAAGTCA 1351
      ||::: ||::: ||::: ||::: ||:::
1768 .....AsnThr.....ArgMetaspMetgluasngluc 1778
1352 TCCAGTGTATAT..... 1364
      ::::: |||
1778 r1met1euser1yrgluc1yasp1ygl1u1alaser1sg1yleu1a 1795
1365 .....TTGAATTAAGGCGCTACAGTTCTTANGCAGCCGAT..... 1400
      ::::: ||::: ||::: ||::: ||:::
1795 spserasn1leser1yrg1yser1yrg1uc1u1p1ro1as1p1ro1ys1 1811
1401 ...TATGACTCCACATTTGCAATATC..... 1424
      ||::: ||::: ||::: ||::: ||:::
1812 ThrGlnaspThrserpheser1le1g1y1yrg1u1alser1u1 1828
1424 ..... 1424
1828 uclugluasp1ugluc1ugluc1u1nar1gser1y1p1roserVal1 1845
1425 .....ACCAAGATGATTCGATTTAATCTATCA 1454
      ::::: ||::: ||::: ||::: ||:::
1845 erc1u1a1h1s1eusergluas1p1ugluc1u1as1p1er1u1as1p1 1861
1455 ACCATAGGGAAGACTGTATCTTCAGATGAT 1487
      ||| ||| ||| ||| ||| |||
1862 1leal1g1y...Aspserasp1euln1as1p1er1 1871

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-188-582-11

seq_documentation_block:

```

: Sequence 11, Application US/08188582
: Patent No 5534410
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comal, Lucio
: APPLICANT: Dynlact, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Rupert, Siegfried
: APPLICANT: Tanese, Naoko
: APPLICANT: Wang, Edith
: APPLICANT: Weinzierl, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE

```

```

: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRIGHTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/188,582
: FILING DATE: 28-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1893 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-188-582-11

```

```

Alignment_scores:
  Quality: 181.50      Length: 561
  Ratio: 0.717        Gaps: 21
  Percent Similarity: 45.098      Percent Identity: 21.034

```

alignment_block:

US-09-687-230-1 x US-08-188-582-11 ..

```

Align seg 1/1 to: US-08-188-582-11 from: 1 to: 1893
177 AAGAGCCACAGTGGACAAACACCTCTACGAGAGTATGTAGAGAGCC 226
      |||||:::
1375 LysLysArgArgValG1yThrThrValHisCysAspTyrLeuAsnArgPr 1391
227 CTTGAAGCTGGTCCCAAGTAGAGGGAAGCAAGACGACGAACTCTCA 276
      ||| :::::
1391 oh1s1ys1er1le1h1s1ar1g1r1g1Thr1asp1rometVal1Thr1euser 1408
277 CGGGACGCTCGGGGCGACGACCTCCAGCCCTTTCGAAGACAAACGATCAT 326
      ::::: ||::: ||::: ||::: ||:::
1408 er1leu1eul1ser1le1le1asn1as1p1er1ar1g1asp1eul1p1ro1as1n1Thr1y 1424
327 GACAAACACAGACACGAAAGCGGAAAGAAAGACAAAGGAGAGAGAGA 376
      ||| ||| ::::: |||
1425 Pro1p1e1h1s1Thr1Pro1Val1asn1al1ys1Val1Val1Lys1asp1Tyr1Tyr1 1441
377 GATTCAGGGAAGAAAGGGAAGAAACGAGAAAGTTAAGAGAGATA 426
      ||| ::::: ||::: ||::: ||::: ||:::
1441 e1le.....Thr1Arg1Pro1met1asp1eul1n1Thr1leu1al1g1u1as1n1V 1455
427 AAAAGAACGA.....GATCGAGACCG..... 449
      ::::: ||::: ||::: ||::: ||:::
1455 al1Arg1ys1ar1g1eul1y1r1Pro1ser1ar1g1u1gl1u1p1he1ar1g1u1h1s1eul1 1471
450 .....GTGAGAAATGAGCGACAGAAAGATCTCCAGTGCACGCCCTCT 493
      |||::: ||::: ||::: ||::: ||:::
1472 Leu1leu1Val1Lys1asn1er1al1a1Thr1Tyr1asn1g1y1p1ro1Lys1h1s1er1eul1h 1488
494 GAGA.....TTAGACTTGCCCTCCTGAGGAAGCCCTTCACAA 528

```



```

1488  ....
1489  rgnllseerGlserMetleuaspLeuCyaspGluLys..... 1501
529  GCTCTTTAGCAAAAGAAAGTAGACAGACACCCCTTCAAGAGCT 578
1502  ....
579  TTGAATCAACATGATGAGACAAATTGCAGAGAAAAGATCCAAAGCTTTC.. 626
1516  lIeAsnProLeu.....LeuAspAspAspGlnValAlaIaphese 1529
626  ....
1529  rPheIlleuAspAsnIleValThrGlnLysMetMetAlaValProAsps 1546
627  ....
1546  eRTPrProPheHISHisProValAsnLysLysPheValProAspLysTyr 1562
669  ATGATCATTTAAACACCCCAATGATTTTATGACATGAAGAAAAGATCAA 718
1563  LysValIleValAsnProMetAspLeuGluThrIleArgLysAsnIleSe 1579
719  GAACATGACATACAGTCCATAGACAACTAAAGATTAAGTCAACTAA 768
1579  rLysHISLysTyrGlnSerArgLysSerPheLeuAspAspValAsnLeu 1596
769  TGTGTACTAATGCCATGATTTTAAATTAACAGACACATTTATTAATAA 818
1596  lIeLysAlaAsnSerValLysTyrAsnGlnProGlnSerGlnTyrThrLys 1612
819  GGTGCAAGAAGCTGTTCACACAGATGAAGAAATTTCTTACGAC... 863
1613  ThrAlaGlnIleValIleValAsnValCysTyrGlnThrLeuThrGluTyrAs 1629
864  .GAAGATTCAGAGCCTGAG... 884
1629  pGlnHISLeuThrGlnLeuGlnLysAspIleCysTyrAlaLysGluAla 1646
885  ....
1646  lAlaLeuGlnIleValIleGlnLeuGlnSerLeuAspProMetThrProGlnPro 1662
915  AAACCTGGAAGCAAGAAATGGAACAGACACCTCACAGCT..... 956
1663  TyrThrProGlnProAspLeuTyrAspThrAsnThrSerLeuSerme 1679
957  .GGGAGGAGGAGGCTGCTGGCAGAGAGAGAGAGAGACTCT...G 999
1679  tSerArgAspAlaSerValPheGlnAspGlnSerAsnMetSerValLeuA 1696
1000  GAGATGCCAAGACACAGCCTTCAAGAGTCCAGCAAGAAATTAATAAG 1049
1696  spLleProSerAlaThrProGlnLysGlnValThrGlnGlnGluLysp 1712
1050  AAAGACAAAGAT..... 1061
1713  GlysAspLysPheLeuAlaAspGlnGluGlnGlyThrValGlnGlnProG 1729
1062  ....
1729  nAlaSerValLeuTyrGlnLysPheLeuMetSer.....GlnGlyG 1743
1102  AGCAGAGACAGCTTGAACCGCATGCTGAAGGAATCTGGAGGAAGCTGACC 1151
1743  LysAspAspGlnGluAspAlaGlySerAspGlnGluGlnLysAspAsn... 1757
1152  AGCGGCTTGTGAACACTCAGTGGCAATTTGAAGAAGAAACCAAGATGG 1201
1758  ....
1202  AACACAGACGTTGGACTTCTCCATCTGTCGATCCCATTTGTAGAGACG 1251

```

```

1771  rAspValGlySerGlyGlyIleArgProLysGlnProArgMet..... 1785
1252  CAGCTACTGCTGCTGTGAGACTGGGATGACAACTGGAAGCTTCAGTCT 1301
1786  ....
1302  GGAGTAAATACCTTTGCCAGGGGTTCAAGAGAGATTAAGCAACAAAGTAC 1351
1789  ....
1352  .TCCAGTGTATAT..... 1364
1799  rMetMetSerTyrGlnGlyAspGlyGlnGluAlaSerHISGlyLeuGlnA 1816
1365  ....
1816  spSerAsnIleSerTyrGlySerTyrGlnGlnLysProAspProLysSerAsn 1832
1401  ...TATGCTCCACATTTGCCAAATATC..... 1424
1833  ThrGlnAspThrSerPheSerIleGlyGlyTyrGlnValSerGlnGlu 1849
1424  .... 1424
1849  uGlnGluAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1866
1425  ....
1866  eArgGlnValHISLeuSerGlnAspGlnGlnGlnGlnGlnGlnGlnGlnGln 1882
1455  ACCTATGGGAGACCTCTGATCTTCCAAAGTGAAT 1487
1883  lIleAlaGly...AspSerAspLeuAspSerAsp 1892
seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:US-08-646-715-11
seq_documentation_block:
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627

```

```

REFERENCE/DOCKET NUMBER: A-57650-2/ALT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1893 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-11

alignment_scores:
Quality: 181.50      Length: 561
Ratio: 0.717         Gaps: 21
Percent Similarity: 45.098   Percent Identity: 21.034

alignment_block:
US-09-687-230-1 x US-08-646-715-11 ..
Align seg 1/1 to: US-08-646-715-11 from: 1 to: 1893

177  AAGAGCAACAACTCGGACAAACACCTCTACGAGAGTATGTAGAAAGCC 226
|||||.....:.....:.....:.....:.....:
1375  LysLysAspArgValGlyThrThrValHisCysAspTyrLeuAsnArgPr 1391
227  CTTGAGCTGTGCTCTCAAGTAGAGAGGACGACGACCTCCAGCACTCCA 276
|  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1391  HisLysSerLeuHisArgArgArgThrAspProMetValThrLeuSer 1408
277  CGGGAGCTCGGGGACGACGACCTCTCCAGCTTGAAGACAAAGATCATC 326
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1408  erLeuLeuGlnSerLeuLeuAsnAspMetArgAspLeuProAsnThrTyr 1424
327  GACAAACACAAAGGACAGAAAGCGGAAAGAAAGAAAGGAGAGAGCA 376
|||||.....:.....:.....:.....:.....:
1425  ProPheHisThrProValAsnAlaLysValValLysAspTyrTyrLys 1441
377  GATTCACAGCGGAAAGGAGGAGGAGAAACGAGAGAGTTAAGAGAGATA 426
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1441  eLle.....ThrArgProMetAspLeuGlnThrLeuArgGluAsnV 1455
427  AAAAGACCGA.....GATCGAGACCGG..... 449
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1455  AlaArgLysArgLeuTyrProSerArgGluGluPheArgGluHisLeu 1471
450  .....GTGGGAATGAGGACAGAAAGATCTCCAGTGCAGCCCTCTGT 493
|||||.....:.....:.....:.....:.....:
1472  LeuLeuValLysAsnSerAlaThrTyrAsnGlyProLysHisSerLeu 1488
494  GAGA.....TTAGACTTCCTCCCTGAGAAAGCTCTCACAA 528
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1488  rGlnLeuSerGlnSerMetLeuAspLeuLysAspLys..... 1501
529  GCTCTTTAGCCCAACAGAGAGTAGAAGACAGACCCCTTCAGAGAGCT 578
|||||.....:.....:.....:.....:.....:
1502  .....LeuLysGluLysGluAspLysLeuAlaArgLeuGluLysAla 1515
579  TTGAATCAACTGATGAGACAATTGCAGAGAAAGATCCAGTGCCTTC.. 626
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1516  IleAsnProLeu.....LeuAspAspAspGlnValAlaPheSe 1529
626  ..... 626
1529  rPheIleLeuAspAsnIleValThrGlnLysMetAlaValProAspS 1546
627  .....TTTTCATTTCCTGTGACTGATTTTATGCTCTGGCTAGCTCC 668
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1546  erTrpProPheHisHisProValAsnLysLysPheValProAspLys 1562
669  ATGATCATTAACACCAATGATTTTAGTACATGAAGAAAGAAATCA 718

```

```

1563  LysValIleValAsnProMetAspLeuGluThrIleArgLysAsnIleSe 1579
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
719  GAACAATGACTACATCATGATCAAGAACTAAAGAACTACTCAACTCA 768
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1579  rLysHisLysTyrGlnSerArgGluSerPheLeuAspAspValAsnLeu 1596
769  TGTGTACTAATGCCATGATTACAAATTAACAGAGACCTATTATTATAA 818
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1596  leuAlaAsnSerValLysTyrAsnGlyProGlnSerGlnThrLys 1612
819  GCTGCAAAAGAACTGTGCATCAGCAATGAATAATCTTAGCCAG.... 863
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1613  ThrAlaGlnGluIleValAsnValCysTyrGlnThrLeuThrGluTyr 1629
864  .GAAGAAATTCAGACCTGAG..... 884
|||.....:.....:.....:.....:.....:
1629  pGluHisLeuThrGlnLeuGluLysAspIleCysThrAlaLysGluAla 1646
885  .....CAGACATAGACTTCATGGCTGACTTGCGAG 914
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1646  lAlaGluGluAlaGluLeuGluSerLeuAspProMetThrProGlyPro 1662
915  AAACCTCGAAAGCAGAAAGATGGAACAGACACCTCAGACAGT..... 956
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1663  TyrThrProGlnProProAspLeuTyrAspThrAsnThrSerLeuSer 1679
957  .CGGAGACGAGGAGCTGCTGGCAGAGAGAGAGAGAGACTCT.....G 999
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1679  tSerArgAspAlaSerValPheGlnAspGluSerAsnMetSerValLeu 1696
1000  GAGATGCCAGACACACGCTTCAAGAGTCCACGCAAAAGAAATTAAG 1049
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1696  spIleProSerAlaThrProGluLysGlnValThrGlnGluLysLysP 1712
1050  AAAGCAAAAGAT..... 1061
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1713  GlyAspGlyAspLeuAlaAspGluGluGlyThrValGlnGlnProGln 1729
1062  .....ATGCTTGAAGATAGTTTAAAGCAATAATTTAGAGAGAG 1101
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1729  nAlaSerValLeuTyrGluAspLeuLeuMetSer.....GluGlyG 1743
1102  AGCAGAGACGCTTGACCGCATCTGAAAGAAATCTGAGAGAAAGCTGAC 1151
|||.....:.....:.....:.....:.....:
1743  LuAspAspGluGluAspLysArgGluGluGlyAspAsn..... 1757
1152  AGCGGCTTGTGACAGTCAGTGCAGATTTGAAGAAAGAAACGAGATGC 1201
|||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1738  .....ProPheSerAlaIleGlnLeuSerGluSerGlySerAspSe 1771
1202  AACAAAGACGTTGGGACTTTCATCTCTGTGATCCCATTTGAGAGAGAC 1251
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1771  rAspValGlySerGlyGlyIleArgProLysGlnProArgMet..... 1785
1252  CAGGCTACTGCTGTGAGACTGGGAATGACAACTGGAAGACTTAGCT 1301
|||.....:.....:.....:.....:.....:
1786  .....LeuGlnGlu 1788
1302  GGAGTGAATACTTTCAGAGGCTTCAAGAGATTAAGAGAACAAAGTCAC 1351
|||.....:.....:.....:.....:.....:
1789  .....AsnThr.....ArgMetAspMetGluAsnGluGluSe 1799
1352  TCCAGTGTATAT..... 1364
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1799  rMetMetSerTyrGluGlyAspGlyGlyGluAlaSerHisGlyLeuGlu 1816
1365  .....TTGAATATAGGCCCTACAGTTCTTATGACACCGCAT..... 1400
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1816  spSerAsnIleSerTyrGlySerTyrGluGluProAspProLysSerAsn 1832
1401  ...TATGACTCCACATTTGCAATATC..... 1424
|||.....:.....:.....:.....:.....:

```



```

883 GTTGCACCTAGGAAATGTTCTTAGCCAGAAAGATTTCAGAGCCCTGA 882
571 .....:|||||:
883 AGCAGAGCATAGACTTCATGGCTGACTTCAGAAATCTGAAAGAGGAAA 932
579 .....:|||||:
579 LUGLUSPIEUGLUTHTLEUYSERLEUGLUSGLUSNLYSARGALIA 595
933 ...GATGGAACAGACCTCAGAGAGTGGGAGAGCGAGGCTGCTGCA 979
596 ILEYSGLYCYSGLYSERGLUTHTSERARGLYSARGLYCS.... 610
980 GAGAGAGAGAGAGACTCTGGAGATGCCAAGACACGCCCTTCAGAGATC 1029
611 .....:|||||:
611 ARGLEUYSERPROTHGLYSGLUSPHTHTHTLEUGLINSERL 627
1030 CCAGCAAAATATAAAGAAAGAACAAATATGCTTGAGATAGTTT 1079
627 .....:|||||:
627 EUGLINSGLUSNGLINGLULEUMETLYSERLEU..... 638
1080 AAAGCAATATTTAGAG.....AGAGACGAGAGCA 1111
639 ...GLUGLYASNLEUGLUTHTPHELEUPHEPROGLYTHRGILUASGLIN 654
1112 GCTTACCCGATCTGAGAGAACTGGAGAAAGCTGACCGAGCGCTTG 1161
654 .....:|||||:
654 ULEUVALSERLEUGLINSGLUSNLEUGLUSERLEUTHR..... 667
1162 TGAAACATGCTGCGAATTTGAAAGAGAAACAGATGAAACAGAGC 1211
668 .....:|||||:
1212 TTGGACTTCTCCATCTGTGATCCCATTTGAGAGAGCCAGGCTACG 1261
682 VALGLYASPGILUGLUALALEUARGPHEUTHRLYSGLU..... 694
1262 CCGTGTGAGACTGGGATGACAACTGGAAGACTTCAGTCTGAGTGAATA 1311
694 .....:|||||:
1312 CTTTGACGGGTTCAAGAGATAAAGAACAAAGTACCTCAGTGTTA 1361
694 .....:|||||:
1362 TATTGAATATGGCCCTACAGTCTTATGACCCGATATGACTCCAC 1411
695 .....:|||||:
695 .....ASNGINGLUPROLEUARGLYSERLEUGLUSASNLYSGLUAL 709
1412 ATTTCGAATATACAGCAAGATGATTGATTATTCATTCACACTATG 1461
709 .....:|||||:
709 APHEARGSERLEUGLUSGLUSASNGINGLUPROLEUYSSTHTLEUGLUG 726
1462 GGAAGACTGTGATCTT.....CCAAAGTATTTCAGCATCATGATTT 1505
726 .....:|||||:
726 LUGLUSASPGILINSERLEVALARGPROLEUGLUTHTGLUSASNLYSER 742
1506 TTGGCCAGCTGCCAAGATTATCCGATATGTCAGGCAGATAGTTTACTGA 1555
743 .....:|||||:
743 LEUARGLYSERLEUGLUGLUN..... 749
1556 TGTTTTAAACAAGAGAGGCAATCCAGACCTTACAAGAGATGAGATGT 1605
750 .....:|||||:
750 .....ASPGILINLUTHTLEUARGLYTHHTLEUGLUN.... 758
1606 CATTCGCTGAAGATGAGGCACTATAGACACTTACACAGGAAAGAA 1655
759 .....:|||||:
759 .....LYSGILUTHTGLINLARGARGYSERLEUGLY.....GLU 770
1656 ATGAGAGCATTTAGAGAGACCAAGCGCTTTGGACTCCAGTAC 1705
771 .....:|||||:
771 GLINASPGILINMETHTHTHTHTHTHTHTHTHTHTHTHTHTHTHT 783
1706 TCAAGACAGGCTCATAGCGCTGAAAGCAGTA...ACAAATTTTGGCGTT 1752

```

```

784 .....:|||||:
784 .....LEUGLUPROLEUYSERLEUASPGILINLUALAARGP 797
1753 CAGTTGAGATTTTGTACTCGAAGAGCTGAAATATTCGAAAGAACTT 1802
797 .....:|||||:
797 .....ASNGINGLUPROLEUARGLYSERLEUGLUSASNLYSER 809
1803 GATGAG.....ACCACGATTTGCTGAG 1825
810 LYSGLUGLUSERVALGLUALVALYSSERLEUGLUTHTHTLEUGL 826
1826 GGAATTCAGAGAGCCAGAGATGACCTTTGAGACCC...AGACCCCTG 1872
826 .....:|||||:
826 USERLEUYSERVALGLYGLINLUSNLEUGLUTHTHTLEUYSERPROG 843
1873 GGAACATGATCTCTCTGCGCCCTCATCAGAGAAATGATCATCT 1917
843 LUTHTGLINALPROLEU.THTHTHTHTHTHTHTHTHTHTHTHTHTHT 857

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-687-080-51

seq_documentation_block:
; Sequence 51, Application US/08687080
; Patent No. 5963427
; GENERAL INFORMATION:
; APPLICANT: Gregory Doljanov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
; US-08-687-080-51

alignment_scores:
Quality: 154.50 Length: 750
Ratio: 0.456 Gaps: 29
Percent Similarity: 45.200 Percent Identity: 17.867

```

alignment block:

US-09-687-230-1 x US-08-687-080-51 ..

Align seg 1/1 to: US-08-687-080-51 from: 1 to: 1312

```

153 CGGTGCGACATGGCGCAAGACACACAGACACAAAGTCGCAACACACCT 202
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
196 ArgInlThrGlnGlnGlnValValGlnGlnGlnGlnGlnGlnGlnGlnGln 212
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
203 CTACGAGAGATGTAGAGAGACCCCTGAAGCTG..... 236
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
212 rLeuLysGlnThrTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 229
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
237 ..... 254
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
229 eLysGlnAlaGlnLeuThrSerSerLysGlnLeuValLysSerTrpGln 245
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
255 AACGAGTCACCCGAACTCTCCAGC.....GCGAGCTCGGGCGACGACTC 298
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
246 AsnGlnLeuAspProLeuLysAsnArgLeuLysGlnGlnGlnGlnGlnGln 262
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
299 CAGCCTCTCGAAGACAAACAGATCATGACAAACACAGACAGACAAAGC 348
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
262 uSerLysIleMetLysLeuAspAsnGlnIleLysAlaLeuAspSerArgL 279
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
349 GGAAGAAAGACAAAGAAAGACAGACAGATTCAGGGGAAAGAAAGGG 398
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
279 yLysGlnMetGlnLysAspAsnSerGlnLeu.....GlnGlnLysMet 293
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
399 AGAAGA..... 404
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
294 GlnLysValPheGlnGlnThrAspGlnGlnLeuAsnAspLeuTrpHisAs 310
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
405 ....CGGAGAGAGCTTAAGAGATAAAGAAAG.....C 435
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
310 nHisGlnArgThrValArgGlnLysGlnArgLysLeuValAspGlnHisA 327
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
436 GAGATCAGACCGGGGTGAGACATGAGCA.....GAAAA 470
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
327 rGlnGlnLeuGlnLysLeuAsnLysGlnSerArgLeuLeuAsnGlnGlnLys 343
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
471 GATCTCCAGTGTACGCCCTGTGAGATTGACTTGCTCTCTGTGAGAG.. 518
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
344 SerGlnLeuLeuValGlnGlnGlnArgLeuGlnLeuAlaAspArgHis 360
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
519 .....CCTCTCAGAGCTCTTACAGCAAC 543
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
360 sGlnGlnHisIleArgAlaArgAspSerLeuIleGlnSerLeuAlaThrG 377
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
544 AAGAGAGAGTA.....GAGACAGACCCCTTCAAGAGCT.... 578
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
377 InLeuGlnLeuAspGlyPheGlnArgGlyProPheSerGlnArgGlnIle 393
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
579 .....TTGATCAACTGATGAGACAAATTCAGAGAAAGATCCAAAGTGC 622
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
394 LysAsnPheHisLysLeuValArgGlnArgGlnGln..... 405
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
623 TTTCTTTTCAATTCCTGTGACTGATTTTATGCTCTGCTGCTACTCCATGA 672
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
406 .....GlyGlnAlaLysT 410
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
673 TCATTAACACCCAAATGATTTAGTACATGAAAGAAAGATCAAGAAC 722
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
410 hAlaAsnGlnLeuMetAsnAspPheAlaGlnLysGlnThrLeuLysGln 426
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
723 AATGACTATAGTCCATAGAGAACTAAAGATTAACCTC...AATCTAAT 769
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
427 LysGlnIleAspGlnIleArgAspLysLysThrGlyLeuGlnArgIleI 443
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
770 GTGCTACTAATGCAATGATTTACATAAACCAGAG.....ACCA 807
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
443 eGlnLeuLysSerGlnIleLeuSerLysLysGlnAsnGlnLeuLysAsnV 460
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:

```

```

808 TTTATTATATAA.....GCTGCAAGAGACTGTGTGCAC 839
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
460 allYsTrpGlnLeuGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 476
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
840 TCAGGAATGAAAATTCTTACCCAGGAAGA..... 869
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
477 LeuAspGlnGlnLeuLeuLysAlaGlnArgGlnLeuSerLysAlaGln 493
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
870 .....ATTGAGCCTGAAGCAGACATA.....GACT 897
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
493 sAsnSerAsnValGlnThrLeuLysMetGlnValIleSerLeuGlnAsnG 510
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
510 IuLysAlaAspLeuAspArgThrLeuArgLysLeuAspGlnGlnMetGln 526
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
948 TCACAGAGTGGGACGACGAGCGCTGCTGCGACAGAGAC.....AGACA 991
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
527 GlnLeuAsnHisIleThrThrThrArgThrGlnMetGlnMetLeuThrL 543
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
992 GCACTGTGAGATGCCGACACAGCAGCGCTTCAAGAGTCCAGCAAGAA 1040
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
543 sAspLysAlaAspLysAspGlnGlnIleArgLysIleLysSerArgHis 560
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1041 .....AATAAAAGAAAGAC 1055
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
560 eAspGlnLeuThrThrSerLeuLeuGlyTrpPheProAsnLysGlnLeu 576
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1056 AAGATATATGCTTGAAGATTAAGTTTAA..... 1082
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
577 GlnAspTrpLeuHisSerLysSerLysGlnIleAsnGlnThrArgAspA 593
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1082 ..... 1082
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
593 gLeuAlaLysLeuAsnLysGlnLeuAlaSerSerGlnGlnAsnLysAsnH 610
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1083 .....AGCAATAATTTAGACAGACAGACAGACAGCTT..... 1115
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
610 IsIleAsnAsnGlnLeuLysArgLysGlnGlnGlnLeuSerSerTrpGln 626
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1116 GACCGATCGTGAAGAAATCTGAGAGAAAGCTGACAGCGCTTGTGAA 1165
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
627 AspLysLeuPheAspValLysGlnSerGln.....As 637
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1166 CACTCAGTGGCAATTTGAAGAGAAAGAACCAAGAT..... 1199
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
637 pPheGlnSerAspLeuAspArgLeuLysGlnGlnIleGlnLysSerSerL 654
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1200 .....GGAACAACGACGTTGGGACTTCTCCATCCT 1229
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
654 yAsnArgAlaMetLeuAlaGlyAlaThrAlaVal...TyrSerGlnPhe 669
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1230 GTGCAATCCCATGTAGAGAGCCAGCGCTACTGC.....CTGCTGAG 1270
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
670 IleThrGlnLeuThrAspGlnAsnGlnSerCysCysProValCysGlnArg 686
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1271 ACTGGAATGACAACAGTGAAGAACTCAGTCTGAGAGGAATCTTGACAG 1320
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
686 gValPheGlnThrGlnAlaGlnLeuGlnGlnValIleSerAspLeuGlnS 703
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1321 GG..... 1322
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
703 eLysLeuArgLeuAlaProAspLysLeuLysSerThrGlnSerGlnLeu 719
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1323 ....TTCAAGAGAGATAAAGAAACAA.....GTCACCTCCAGTGT 1360
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
720 LysLysLysGlnLysArgArgAspGlnMetLeuGlnLeuValProMetArg 736
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1361 ATATTGAATTATAGGCCCTACAGTCTTATGACGACGCAATTGACTCCA 1410
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
736 gGlnSerIleIleAspLeuLysGlnLysGlnIleProGlnLeuArgAsnL 753
    |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:

```

```

1411 CATTGCAATATACGACGAGATGATTCGATTTATCTATTCACCTAT 1460
      |||:|||||:|||||: |||
753 yslenglnasnvalasnargaspilleglnhrgleuylasnasprlleu 769
      |||:|||||:|||||: |||
1461 GGGGAGAGCTGATCTTCAGATGATTCAGATTCAGATGATTTGGC 1510
      |||:|||||:|||||: |||
770 GlnGlnGlnThrLeuLeuGlnGlnThrLeuMetProGlnGlnGlnSerAla 786
      |||:|||||:|||||: |||
1511 CACGTGCCAAGATTTCCGTATGTCATGGCAGATAGTTACTAGGTGTT 1560
      |||:|||||:|||||: |||
786 svalys.....LeuThrAspValT 793
      |||:|||||:|||||: |||
1561 TACAAAGAGAGGCGATTCAGACCTACAGAGATGAGATGTCATTC 1610
      |||:|||||:|||||: |||
793 hllleMetGlnArgPheGlnMetGlnLeuLysAspValGlnArgLysIle 809
      |||:|||||:|||||: |||
1611 CCTGAGATGAAAGCCATCTAGGACACTTGACACAGGAAAGAAATGGA 1660
      |||:|||||:|||||: |||
810 AlagInGlnAlaAlaLysLeuGlnGlnLysLeuAspArgThrValG 826
      |||:|||||:|||||: |||
1661 GCAGAT..... 1667
      |||:|||||:|||||: |||
826 nllnValasnGlnGlnLysGlnGlnLysGlnHisLysLeuAspThrValS 843
      |||:|||||:|||||: |||
1668 ..ACAGAGTAGAGCCACAGGCGCTTGGACTCCAGTACATCAGACAGG 1715
      |||:|||||:|||||: |||
843 exSerLysIleGlnLeuLysAsnArgLysLeuIleGlnAspGlnGlnGln 859
      |||:|||||:|||||: |||
1716 CTCATAGCGCTGAAGCAGTACAAATTTGGCGTCCAGTTGAAGTTT 1765
      |||:|||||:|||||: |||
860 lIleGlnHisLeuLysSerThrThrAsn..... 868
      |||:|||||:|||||: |||
1766 TCACTCTGAAAGAGCTGAATATTCAGAGAAAGAACTGTGAGACACCA 1815
      |||:|||||:|||||: |||
869 .....GlnLeuLysSerLysLeuGlnLysLeuIleSerThrA 880
      |||:|||||:|||||: |||
1816 GATTCCTCAGGAGACTCAG...GAAGCCAGATGAAACGTTTGAACACC 1862
      |||:|||||:|||||: |||
880 snleGlnArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 896
      |||:|||||:|||||: |||
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.us-08-592-126-148

seq_documentation_block:
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; City: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960

```

```

; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
; US-08-592-126-148

alignment_scores:
  Quality: 153.50      Length: 750
  Ratio: 0.453        Gaps: 29
  Percent Similarity: 45.200  Percent Identity: 17.867

alignment_block:
US-09-687-230-1 x US-08-592-126-148 ..

Align seg 1/1 to: US-08-592-126-148 from: 1 to: 1312

153 CGGTGGACATGGGCAAGACACAGCAAGCAAGTCGGCAAAACACT 202
      |||:|||||:|||||: |||:|||||:|||||: |||
196 ArgGlnThrGlnGlnGlnValLysGlnLysGlnMetGlnLeuLysTY 212
      |||:|||||:|||||: |||
203 CTACGAGAGATGATGAGAGAGCCCTTGAGCTG..... 236
      |||:|||||:|||||: |||
212 rleuLysGlnThrLysGlnLysLysAlaCysGlnLysLeuArgAspGlnIleThrS 229
      |||:|||||:|||||: |||
237 .....GTCTCAAGTGAAGAGCG 254
      |||:|||||:|||||: |||
229 erLysGlnAlaGlnLeuThrSerSerLysGlnLysValLysSerLysGlu 245
      |||:|||||:|||||: |||
255 AACGAAAGTCCAGACCTCTCCAG.....GGACGCTGGGGCAGACTC 298
      |||:|||||:|||||: |||
246 AsnGlnLeuAspProLeuLysAsnArgLysGlnLysGlnHisAsnIle 262
      |||:|||||:|||||: |||
239 CAGCCTCTCGAAGCAAAACGATCATGACAAACCAAGGACAAAGC 348
      |||:|||||:|||||: |||
262 uSerLysIleMetLysLeuAspAsnGlnLysAlaLeuAspSerArgL 279
      |||:|||||:|||||: |||
349 GGAAGAAAGCAAGAAAGAGAGAGACAGATTCAGCGGGAAGAAAGGG 398
      |||:|||||:|||||: |||
279 yslYsGlnMetGlnLysAspAsnSerGlnLeu.....GlnGlnLysMet 293
      |||:|||||:|||||: |||
399 AGAAA..... 404
      |||:|||||:|||||: |||
294 GlnLysValPheGlnGlnGlnThrAspGlnGlnLeuAsnAspLeuThrHisAs 310
      |||:|||||:|||||: |||
405 ...CGAGAGAGATTAAAGGAGATAAAAAGAG.....C 435
      |||:|||||:|||||: |||
310 nHisGlnArgThrValArgGlnLysGlnArgLysLeuValAspCysHisA 327
      |||:|||||:|||||: |||
436 GAGATCGAGACCGGCGTGGAGATGAGCA.....GAAAA 470
      |||:|||||:|||||: |||
327 rglLysGlnGlnLysLeuAsnLysGlnSerArgLysLeuAsnGlnLys 343
      |||:|||||:|||||: |||
471 GATCTCAGATGTACAGCCCTGTGAGATTGACCTCTGAGAG.. 518
      |||:|||||:|||||: |||
344 SerGlnLeuLeuValGlnGlnGlnArgLysGlnGlnGlnAlaAspArgHi 360
      |||:|||||:|||||: |||
519 .....CTCTCAAGCTCTTAGCCAAAC 543
      |||:|||||:|||||: |||
360 sGlnGlnHisIleArgAlaArgAspSerLeuIleGlnSerLeuAlaTrng 377
      |||:|||||:|||||: |||
544 AAGAGAGATA.....GACAGACACCCCTTCAGAACT..... 578
      |||:|||||:|||||: |||
377 lLysGlnLysLeuAspLysPheGlnArgGlyProPheSerGlnArgGlnIle 393
      |||:|||||:|||||: |||
579 .....TTGAATCAACTGATGAGACAATTGACAGAAAAAGATCAAGTGC 622
      |||:|||||:|||||: |||
394 LysAsnPheHisLysLeuValArgGlnArgGlnGln..... 405

```

```

623 TTTCCTTTCATTTCTGTGACGATTTTATTTCTGCTACTCCATGTA 672
    |||
406 .....GlyGluAlaLysr 410
673 TCATTAAACCCCAATGGATTTCATACATGAAGAAGAAATCAAGAAC 722
    |||
410 hrlaasnglnleuMetIasnppheAlaGluLysGlnThrLeuLysGln 426
723 AATGACATACAGTCATAGAAAGCTAAAGCATTAACCTTC...AAGCAATAT 769
427 LysGlnIleAspGlnIleArgAspLysLysThrGlyLeuLysArgIleI 443
770 GTGTACTAATGCATGATTATTAATAAACAGAG.....ACCA 807
443 egluleuLysSerGlnIleuSerLysGlnAsnGlnLeuLysAsnV 460
808 TTTATTTATAA.....GCTGCAAGAAAGCTGTGCAC 839
    |||
460 allYsTYrGlnLeuGlnGlnLeuGlnLysSerAspArgIleLeuGln 476
840 TCAGAAATGAAATTCCTTACCCAGCAAGAA..... 869
477 LeuAspGlnIleuLysAlaGlnArgGlnLeuSerLysAlaGlnLys 493
870 .....ATTCAGACCTGAAGCAGACCATR.....GACT 897
493 sasnSerAsnValGlnThrLeuLysMetGlnValIleSerLeuGlnAsn 510
898 TCATGCTGCTGCGAGAAACTCGAAGACGAAGAAATGGAACAGACACC 947
510 LuLysAlaAspLeuAspArgThrLeuArgLysLeuAspGlnLysMetGln 526
948 TCACAGAGCTGGGAGCAGCAGCCTGCGAGAGAGAG.....AGAGA 991
527 GlnLeuAsnHisThrThrThrArgThrGlnMetGlnLysLeuThrLys 543
992 GGACTGTGAGATGCGAGACACAGCCTTCAGAGTCCAGCAAGAA. 1040
543 sAspLysAlaAspLysAspGlnIleArgLysIleLysSerArgHis 560
1041 .....AATMAAAGAAAGAC 1055
560 eAspGlnLeuThrSerLeuGlnLysTyrrPheProAsnLysLysGlnLeu 576
1056 AAGATATGCTTGAAGATTAAGTTTAA..... 1082
577 GlnAspTrpLeuHisSerLysSerLysGlnLysAsnGlnThrArgAspArg 593
1082 ..... 1082
593 gLeuAlaLysLeuAsnLysGlnLeuAlaSerSerGlnGlnAsnLysAsnH 610
1083 .....AGCAATTAATTTAGAGAGAGCAGAGACGTT..... 1115
610 lAsIleAsnAsnGlnLeuLysArgArgGlnGlnLysLeuSerSerTyrrGln 626
1116 GACCGCATCTGTAAGAAATCTGAGAGAAAGCTGACAGCGCGCTTGTA 1165
    |||
627 AspLysLeuPheAspValCysGlnSerLys.....As 637
1166 CAGTCAGTGCATTTGAAAGAAAGAAACCAAT..... 1199
637 pPheGlnLysSerAspLeuAspArgLeuLysGlnGlnLysLeuSerL 654
1200 .....GGAACACAGAGCTGGGACTTCTCCATCT 1229
654 ysglnArgAlaMetLeuAlaGlnLysArgAlaVal...TyrSerGlnPhe 669
1230 GTGATTCCTCATTTGAGAGAGCCAGCTACTGC.....CTGGTAG 1270
    |||
670 lIeThrGlnLeuThrAspGlnAsnGlnLysCysProValCysGlnLys 686

```

```

1271 ACTGGAAATGACAACCTGGAAGACTTCAGTGTGAGAGTGAATACTTGACG 1320
    |||
686 gValPheGlnThrGlnAlaGlnLeuGlnGlnValIleSerAspLeuGln 703
1321 GG..... 1322
703 eLysLysLeuArgLeuAlaProAspLysLeuLysSerThrGlnLysGlnLeu 719
1323 ...TTCAAGAGAGTAAAGAAACAAA.....GTCACTCCAGTGT 1360
    |||
720 LysLysLysGlnLysArgArgAspGlnMetLeuGlnLysLeuValPheMet 736
1361 ATATTGAATTAATGAGCCCTACAGTTCTTATGCACCGCATTAATGACGCCA 1410
736 gGlnSerIleIleAspLeuLysGlnLysGlnLysLeuProGlnLeuArgAsn 753
1411 CATTTGCAATATACAGCAAGAGATTCGATTTAATCTATTAACCTAT 1460
    |||
753 yslLeuGlnAsnValAsnArgAspIleGlnArgLeuLysAsnAspIleGln 769
1461 GGGGAGACCTGTGATCTTCCAGTGAATTCAGCATTCACATGAGTTTGGC 1510
    |||
770 GlnGlnGlnThrLeuLeuGlnLysThrIleMetProGlnLysGlnLysAla 786
1511 CAGCTGCCAAGATTAATCCGTATGTCATGCGAGATAGTTTACTGATGTT 1560
    |||
786 sValCys.....LeuThrAspValT 793
1561 TAACAAAGAGAGGAGCTTCAGACCCCTACAGAAAGATGAGATGTCATG 1610
    |||
793 hrlMetGlnArgPheGlnMetGlnLysAspValGlnArgLysIle 809
1611 CTTGAGATGAGAGCCATTAAGACACTTTCAGACACTTCACAGAGAAAGAAATGGA 1660
    |||
810 AlaGlnIleAlaIleLysLysGlnGlnLysLeuAspArgThrValGln 826
1661 GCAGATT..... 1667
826 nGlnValAsnGlnLysGlnLysGlnHisLysLeuAspThrValS 843
1668 ..ACAGAAATGAGCCACAGAGGCGTTTGCATCCAGTACTCAAGACAGG 1715
    |||
843 eSerLysIleGlnLeuAsnArgLysLeuIleGlnAspGlnGlnGln 859
1716 CTCATAGCGCTGAAGACAGTAACTAATTTGGCGTCCAGTTGAAGTTT 1765
    |||
860 lIleGlnHisLeuLysSerThrThrAsn..... 868
1766 TGACTCTGAAGAGCTGAATATATTCAGAGAAAGCTTGATGAGACACCA 1815
    |||
869 .....GlnLeuLysSerGlnLysLeuGlnIleSerThr 880
1816 GATTCCTCAGGAGACTCCAG...GAAGCCAGAAATGAAGCTTTGAGACAC 1862
    |||
880 snLeuGlnArgArgGlnGlnLysGlnGlnIleValGlnLysSerThr 896

```

seq_name: /cgn2_6/protdata/2/1aa/5a_COMB.pep:US-07-853-913-2

seq_documentation_block:

Sequence 2, Application US/07853913

Patent No. 5338839

GENERAL INFORMATION:

APPLICANT: McKay, Ronald D.G.

INVENTOR: Lendahl, Urban

TITLE OF INVENTION: Nestin Expression As An Indicator Of

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: U.S.A.


```
966 GGAGCTCTGTCAGACAGAGAGA..... 989
    |||
527 GlnGlyProLeuGlnLysGlnThrLeuLysAlaLeuGlyGlnProLe 543
990 .....GAGACTCTGA..... 1001
543 uMetSerLeuLysIleGlnAsnTyrGluThrAlaGlyLysGlnCysA 560
1002 .....GATCCGGAACACAC.....GCCCTCAAGAGTCCACAGAA 1040
560 snSerSerThrGlnGlyHisLeuGlyThrLeuGlnGlyProGlnLysGln 576
1040 ..... 1040
577 LysGlnIleProLeuLysSerLeuGlnGlnLysAsnValGlnSerGlnL 593
1040 ..... 1040
593 sThrLeuGlnAsnGlyValProValLeuSerGlnLeuGlnLysGlnA 610
1041 ..AATAAAAGAAAGACAAAGATATGCTTGAAGATTAAG..... 1076
    |||
610 sPThrArgThrGlnAspGlnGlnLeuMetSerProLysGlyThrLeuLys 626
1077 .....TTTAA 1081
627 ArgPheSerSerLeuGlyLysGlnSerGlnGlnValAlaArgProSer 643
1082 AAGCAATTAATTAGAG..... 1097
643 sGlnGlyAsnLeuGlnSerTrpThrAlaPheLysGlnGlnSerGlnHis 660
1098 .....AGAGACAGACAGACGCTTGAACCCATGCTG... 1127
660 roleuGlyPheProGlyAlaGlnLysGlnMetLeuGlnArgLeuValGln 676
1127 ..... 1127
677 LysGlnLysPglInsPheProArgSerProGlnGlnLysPglInsVal 693
1128 .....AAGAAAT 1134
693 acyArgProLeuGlnLysGlnAsnGlnGlnProLeuGlnLysTrpGln 710
1135 CTGAGAGAAAGCTACACAGCGCGCTTGTGAACAGTCACTGCGAATTGAA 1184
    |||
710 lagGlnGlyGlnIleLeuGlnArgLeuIleGlnLysGlnSerGlnLys 726
1185 AGAGAAACCA..... 1196
727 LeuArgSerProGlnGlnLysPglInsValAlaGlyArgSerLeuGln 743
1197 .GATGAGAACAGACGTTGGACTTCCATCCGTGATCCCAT... 1241
743 sGlnAsnGlnGlnProLeuGlnLysTrpGlnAlaGlnAspGlnMetLeu 760
1241 ..... 1241
760 LysArgLeuIleGlnLysGlnSerGlnLysSerProGlnGln 776
1242 .....GTAGAGAGACAGCGTACTGCTGGTGAAGTGAATGAC 1282
777 AsnGlnArgIleGlyLysPro.....LeuGlnArgGlnAsnGlnLys 790
1283 AACT.....G 1287
790 sSerLeuArgTyrLeuGlnGlnAsnGlnGlnThrPheValProLeuGln 807
1288 GAAGACTTCAGTCTGAGTGAAATCTTGCAGGCTCAAGAGAGATAA 1337
807 ernArgAsnGlnArgProLeuArgSerLeuLysValGlnGlnGlnGln 822

1338 AGAACAAGATCACTCCAGTGTATATTGAATTATGGCCCTACATTC 1387
    |||
823 GlnArgIleValLysProLeuGlnLysValSerGlnAspSerLeuGly 839
1388 TTATGCAACCGCATATATGCTCCACATTTGCCAAATATACAGAGATAT 1436
839 rLeuAlaGlnGlnAsnValGlnProLeuArgTyrLeuGlnGlnAsp 856
1437 .....TCTGATTTAACTATTCACCTATGGGAA 1466
856 ysiLeuAsnLysSerLeuGlnAspLysThrHisLysSerLeuGlySer 872
1467 GACTCTGATCTTCCAGTGAATTCAGCATC..... 1496
873 LeuGlnAspArgAsnGlyAspSerIleIleProGlnGlnSerGlnThr 889
1496 ..... 1496
889 rGlnValSerLeuArgProProGlnGlnLysPglInsArgIleValAsn 906
1497 .....CATGAGTTTTGGCCACGTCGCAAGATTTATCCG 1529
906 lseuGlnLysGlnSerGlnGlnPheSerArgSerSerGlnGlnGln 922
1530 TATGTCATGCGCAGATAGTTACTGATGTTTAACAAAAGAGGCCATTC 1579
923 GlnValMetGlnArgSerLeu.....GlnGlyGlnAsnH 934
1580 CAGAGCCCTACAA.....GAGATGCAGATGTCATTCGCCGAGATGAC 1623
934 sGlnSerLeuSerSerValGlnLysGlnAspGlnMetValGlnSerGln 951
1624 GCCATCTAGACACCTTGACACAGAAAGAAATGACAGATTTACAGAA 1673
951 euGlnLysGlnSerGlnAspSerGlyLysSerLeuGlnAspGlnSerGln 967
1674 GTAGAGCCACAGCGCGCTTGGACTCAGTACTCAAGACGCTCATAC 1723
968 ...GlnThrPheGlyProLeuGlnLysGlnAsnAlaGlnSerLeuArg 983
1724 GCTGAAGACAGTAAACAATTTGGCGTCCAGTTGAAGTTTTCACGCTG 1773
983 rLeuAlaGlyGln.....AspGlnG 990
1774 AAGAAGCTGAATATTTCCAGAGAAATTTGATGAGACACACAGATGCTC 1823
990 LysLys.....GlnLysLeuGlnGlnGlnThrGlnGlnThrLeu 1002
1824 AGGGAATCTCAGAGACCCAGAAATGAACGTTTGACACACAGAC..... 1866
1003 ArgAlaValAlaGlnLysGlnGlnMetAlaValSerProGlnLysValA 1019
1867 .....CCCGTGAACATGAT.....CTGTC 1887
1019 sProGlnLeuProLysProLeuGlnLysAsnAspGlnGlnIleAlaArg 1035
1888 TCTTGGGCTCCCTCATCAGAGA.....ATGCATCTTCTGAAACAAGTGA 1931
1035 rLeuGlnLysGlnAsnGlnGlnSerLeuValSerLeuLysGlnLysGly 1052
1932 CCAATATATCTTAAGAAATTTGACAGCAAGTAACTCCAGGTATATGTA 1981
1052 leuGlnThrValLysSerLeuGlnThrGlnIleIleGlnProLeuGlnThr 1068
1982 GCACGTATGAGATTCGAAAGCAAGTGGGATTTCCATTCCTTCCCGCTC 2031
1069 AlaGlnGlnAspLeuGlnArgArgLysSerIleAspThrGlnGlnProLe 1085
2032 ATGGAACA 2041
1085 uTrpSerThr 1088

seq_name: /cgn2_6/ptodata/2/laa/5B_COMB.pep:us-08-663-112-2
```



```

341 ysllmetcspahnshtlslsgluarqlllelaasnhelylllelupro 357
1134 TCTGGA.....GGAAAGCTGAC 1150
358 ProglyLeuPheArgGlyArgGlyAsnHsProlslsMetGlyMetLeuLly 374
1151 CAGGGGG.....CTTGTAACAGTCAGTCGCAATTGG 1182
374 sarqarqllmetProgluAspIlellelaasn.CysSerLysAspAla 390
1183 AAGAAGAAACACAGATGAGAACACAGAGCTGG.....GA 1217
391 LysValProSerProProProProGlyHsLysTrpLysgluValArgHsAs 407
1218 CTTTCCTCCATCTGTGATCCATTCGTAGAGAGCAGGCTACTGCTGCT 1267
407 pashlysaValThrTrpLeuVal..... 414
1268 GAGACTGGGAATGACAACTGGAGACTTCAGTCGTGAGATGAACTTGC 1317
415 ..SerTrpThrGluHsHsIleGlnGly...SerIleLysTrpIleMetLeu 429
1318 AGGGGTTCAAGAGAGATMAAGAACAAAGTCACTCCAGTGTTAT 1362
430 AsnProSerSerArgIleLysGlyGluLysAspTrpGlnLysTrp 444
seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US93-07261-11

```

```

seq_documentation_block:
  Sequence 11, Application PC/TUS9307261
  GENERAL INFORMATION:
  TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: John H. C. Blasdale
  STREET: One Giralda Farms
  CITY: Madison
  STATE: New Jersey
  COUNTRY: USA
  ZIP: 07940-1000
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  OPERATING SYSTEM: Macintosh 6.0.5
  SOFTWARE: Microsoft Word 5.1a
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: PCT/US93/07261
  FILING DATE: 19930805
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/927,531
  FILING DATE: 07-AUG-1992
  ATTORNEY/AGENT INFORMATION:
  NAME: Blasdale, John H. C.
  REGISTRATION NUMBER: 31,895
  REFERENCE/DOCKET NUMBER: DX0288K
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 201-822-7398
  TELEFAX: 201-822-7039
  INFORMATION FOR SEQ ID NO: 11:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1588 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  PCT-US93-07261-11

```

```

alignment_scores:
  Quality: 147.50
  Ratio: 0.467
  Percent Similarity: 47.094
  Length: 671
  Gaps: 32
  Percent Identity: 19.523

```

```

alignment_block:
  US-09-687-230-1 x PCT-US93-07261-11
  Align seg 1/1 to: PCT-US93-07261-11 from: 1 to: 1588

```

```

309 GAAGACAAAAGAGATGATGACAAACACAGAGAC.....AGAAAGCGGA 352
29 GlnLysLysAsnAspAlaGlnLysAlaLysAspLeuThrLysLysGluSe 45
353 AAGAGAAAAGAAAGAGAGAGAACAGATTCACAGCGGAGGAGAAAAGGGGACA 401
45 rGlnAspSerSerSerSerGlnLysSerLeuLysgluValAsnGlnGluA 62
402 .....AAAGGAGAGAGATTAAAG 419
62 laLeuLysgluLysgluAsnLysgluThrLeuLysLysgluLeuGln 78
420 GAGATTAAGAAAGAGAGATGAGAGACCGGGTGAAGATGAGCAGAAA 469
79 AsnGlnLysgluLysgluGlnLysAsnLysIleLysAspAsnAspGln 95
470 AGATCTCCAGTGTCCAGCCCTGTGAGATTAGAC.....TTGCCCTCGT 513
95 uAlaLeuLysAsnLysgluAsnAspLysAspAspLysLysIleValProl 112
514 AGAAGCTCTTCACAAAGCTCTTTAGCACAACAAAGAAAGTGAACAGACA 563
112 yslLysPro..... 114
564 CCCCTTCAGAGAGCTTTGAATCAACTGATGAGACATTCGACAGAAAAGA 613
115 .....GlnSerValGlnLysAspLeuLysgluMetGlnLeuLysglu 128
614 TCCAAGTCCTTCCTTCATTCCTGTCGACGATTTATTTATTCCTCGCT 663
128 ulys..... 129
664 ACTCCATGATCATTTAAACACCAATGATTTAGTACATGAAAAGAAAG 713
130 ..GlnPheIleLysGlnHsIleLysAspTrpTrpGlnLysgluLys 145
714 ATCAAGAAACATGCTATTCAGTCCATA.....GAAGAACTAAAGATTA 757
146 ArgArgAsnTrpIleLeuArgSerLeuArgArgAspLysLeuArgGln 162
758 CTTCAACTAATGCTGACTAATGACATGATTTACAAATTAACCGAGACCA 807
162 egluGlnLeuGlnLysLeuAsnAlaGlnLeuGlnSerAlaIleAsnGln 179
808 TTTATTATTAAAGCTGCMAAGAG.....CTGTTGACATCA 842
179 euLysgluArgArgAlaSerArgArgProMetValLysMetGlnArg 195
843 GGAATGAATAATTCTTAGCCAGGAAAGAAATTCAGAGCTTACAGACAGCAT 892
196 GlyMetLys.....AspGlnValAspGlnTrpIleLysLysLys 209
893 AGACTTCAGGCTGACTTCAGAAAACCTGAAAGAGAAA..... 932
209 pAspGlnGlnAlaGlnLysAsnGlnLysThrLysAspGlnGlnIleLysAsp 226
933 .....GATGGA.....ACAGACACTCACAGAGTGG..... 959
226 yslGlnAspGlnLysgluGlnIleValGlnThrLysPheTrpGlnMetArg 242
959 ..... 959
243 GluAsnAlaLeuGlygluLeuAspGlnLysgluValArgTrpGlnLysLys 259
960 .....GAGACGAGAGCTGCTGCGACAGAGACAGAGAGAG 993
259 sarqTrpTrpLeuLysgluAspGlnGlygluLysAspLeuLysAspAlaGln 276

```

```

994 ACTCTGAGATGCGAGACACAGCCCTTCAGAGTCCC..... 1031
      :::::  :::::  :::::  :::::  :::::  :::::
276 IuylsleuGlulthrGlyTyrGlyPheArgGlyLysPheProThr 292
      :::::  :::::  :::::  :::::  :::::  :::::
1032 ..... AGCAACAGAAATATAAAGAAAGCAAAAGATATCTTGA 1069
      :::::  :::::  :::::  :::::  :::::  :::::
293 ArgIleLeuValIysArgLysArgAsnLysGluGlnLysLysLeuLysGln 309
      :::::  :::::  :::::  :::::  :::::  :::::
1070 AGTTAGTCTTAAAGCAATATATTTAGAGAGACAGACAGACGCTTGACC 1119
      :::::  :::::  :::::  :::::  :::::  :::::
309 uAspLysGluLysLysLeuIleAlaIleGluGluProAspArgLysL 326
      :::::  :::::  :::::  :::::  :::::  :::::
1120 GCATC...GTGAAGAATCTGGAGAAACCTGACACAGCGCGCTTGTGAAAC 1166
      :::::  :::::  :::::  :::::  :::::  :::::
326 YstIleLysLeuLysAspSerAspLysValValAlaProValAsnLys 342
      :::::  :::::  :::::  :::::  :::::  :::::
1167 AGTCAGTGGCAATTT.....GAAAGAGAAACACAGATGAAACAGCAC 1210
      :::::  :::::  :::::  :::::  :::::  :::::
343 AsnLysSerPheProAspLysPheArgAlaProAspLysLysArgTh 359
      :::::  :::::  :::::  :::::  :::::  :::::
1211 G.....TTGGACTTCTCCATCTGTGATCCCATTTGTAGAG 1248
      :::::  :::::  :::::  :::::  :::::  :::::
359 rMetPheTyrArgLysSerGluLeuPheProIleValProArgLysAsp 376
      :::::  :::::  :::::  :::::  :::::  :::::
1249 AGCCAGGCTACTGCGTGTGAGACTGGGATGACAACGTGAAACATCTCAG 1298
      :::::  :::::  :::::  :::::  :::::  :::::
376 sn.....GluLeuAlaValLysGlyAspSerMetAsp 386
      :::::  :::::  :::::  :::::  :::::  :::::
1299 TCTGAGTGAAT.....ACTTGCAGGGGTTCAAAGA 1330
      :::::  :::::  :::::  :::::  :::::  :::::
387 SerLysValAsnGlyLysLysLeuLysSerThrPheAsnProPheLys 403
      :::::  :::::  :::::  :::::  :::::  :::::
1331 CGATAAAGCAACAAAGTCACCTCCAGTGTATATTTCAATTATGGCGCC 1380
      :::::  :::::  :::::  :::::  :::::  :::::
403 g...ArgArgAsnLysLeuLysGluArgLysMetGlnGluLeuHisLysP 419
      :::::  :::::  :::::  :::::  :::::  :::::
1381 ACAGTCTTATGACACCCATATGACTCCACATTTGCCAATATACGACAG 1430
      :::::  :::::  :::::  :::::  :::::  :::::
419 heLysLysAsnTyrLysLysTyrGlnLysLeuGluArgGluLysArg 435
      :::::  :::::  :::::  :::::  :::::  :::::
1431 GATGAT..... 1436
      :::::  :::::  :::::  :::::  :::::  :::::
436 GluAsnProAspGlyGluProLeuAsnThrProGluIleHisValIle 452
      :::::  :::::  :::::  :::::  :::::  :::::
1437 ...TCTGATTTAATC..... 1448
      :::::  :::::  :::::  :::::  :::::  :::::
452 gProSerAspLeuMetAspLysGlyGluAsnLysSerAlaGlyHisProp 469
      :::::  :::::  :::::  :::::  :::::  :::::
1449 .....TATTCACCTATGGG.....GAAAGACTGTGATCTT 1478
      :::::  :::::  :::::  :::::  :::::  :::::
469 heLysTyrGlnProThrLysGlyLeuLysGluTyrGlnGluSerHisVal 485
      :::::  :::::  :::::  :::::  :::::  :::::
1479 CCAAGTGAATTCAGCATC...CATGAGTTTGGCCACGATGCCAAGATTA 1525
      :::::  :::::  :::::  :::::  :::::  :::::
486 SerLysAspTyrGlnLeuGlnHisGluProProThrLysLeuProGlu 502
      :::::  :::::  :::::  :::::  :::::  :::::
1526 TCCG.....TATGTCATGCCAGATAGTTTACTGGAT..... 1556
      :::::  :::::  :::::  :::::  :::::  :::::
502 rGluLysGluHisValSerArgLysGluIleAspHisGluProProT 519
      :::::  :::::  :::::  :::::  :::::  :::::
1557 .....GTTTACAAAGAGAGGATTCAGGACCCCTACAA... 1592
      :::::  :::::  :::::  :::::  :::::  :::::
519 heLysLeuProGluTyrGluLysGlnHisValSerArgLysGluIle 535
      :::::  :::::  :::::  :::::  :::::  :::::
1593 .....GAGATGAGATGTTCATTTGCGCTGAA...GATGAAGGCCATATAG 1633
      :::::  :::::  :::::  :::::  :::::  :::::
536 AspAsnGluValArgAspGluLeuProGluTyrGluLysGlnHisValSe 552
      :::::  :::::  :::::  :::::  :::::  :::::
1634 G.....ACACTTGAC.....ACAGGAAAGAAATGG 1659
      :::::  :::::  :::::  :::::  :::::  :::::
552 rArgGluTyrGlnLeuAspAsnGlnLysProSerThrLeuLysGluTyr 569
      :::::  :::::  :::::  :::::  :::::  :::::
1660 AGCAGATTACAGAAATTA.....GAGCCACCA 1685
      :::::  :::::  :::::  :::::  :::::  :::::

```

```

569 spGln...ThrGluLeuAlaLysGlyLysAspIleThrAsnLysProHis 584
      :::::  :::::  :::::  :::::  :::::  :::::
1686 GGGCGTTTGGACTCCAGTACTCAAGACAGGCTCATAGCCCTGAAAGCAAT 1735
      :::::  :::::  :::::  :::::  :::::  :::::
585 GluSerValAspGluTyrAspGlnThrGluLeuAlaLysGlyLysAspI 601
      :::::  :::::  :::::  :::::  :::::  :::::
1736 AACAAATTTGGCGTTCAGTTGAAAGTTTTCAGTCTGAAAGACGTGAAA 1785
      :::::  :::::  :::::  :::::  :::::  :::::
601 eThrAsn.....LysProHisGluSerValAspGluTyrAspGlnSer 616
      :::::  :::::  :::::  :::::  :::::  :::::
1786 TATTCAGAAAGAACTTGATGAGACACACAGATTCCTCAGGGAATCCAG 1835
      :::::  :::::  :::::  :::::  :::::  :::::
616 IuLeuAlaLysGlyLysAspIleThrAsnLysProHisGluSerValAsp 632
      :::::  :::::  :::::  :::::  :::::  :::::
1836 GAA.....GCCACAGATGAACGTTTGAGACACAGACC 1867
      :::::  :::::  :::::  :::::  :::::  :::::
633 GluTyrAspGlnThrGluLeuAlaLysGlyLysGluValThrAsnLysP 649
      :::::  :::::  :::::  :::::  :::::  :::::
1868 CCTGGGGAACATG 1880
      :::::  :::::  :::::  :::::  :::::  :::::
649 ohisGluAsnLeu 653
      :::::  :::::  :::::  :::::  :::::  :::::

```

OM of: US-09-687-230-1 to: A_Geneseq_032802:* out_format : pfs

Date: Jul 11, 2002 3:28 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp
-O=/cgn2_1/USPRO_pool/US09687230/runat_11072002_144407_29302/app-query.fasta.1.2401
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DEL0P=6.000 -DEL0EXT=7.000 -START=1 -MATRIX=blsnum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORExt=HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09687230_@CGL_1.203 -NCPU=6 -ICPU=3 -LONGIOG
-DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-687-230-1
Query length: 2307
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 11073796
Search time (sec): 166.540000

```

score_list:
Sequence      Strd Orig      Zscore      EScore Len      Documentation
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAO13504 + 3298.00 5401.12 1.6e-29
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAU16626 + 3275.50 5363.22 1.9e-29
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:AAU37947 + 3073.00 5031.91 6.6e-27
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH83712 + 2013.50 3290.92 9.0e-17
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH83765 + 1716.50 2803.32 1.5e-14
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH40621 + 927.50 1497.44 4.8e-76
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH83835 + 921.50 1487.71 3.7e-75
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAU16815 + 861.50 1401.40 3.3e-70
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH83408 + 738.50 1182.48 1.2e-58
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH85421 + 684.00 1098.48 9.6e-54
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41780 + 476.00 762.49 8.9e-35
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH5881 + 391.00 624.57 5.2e-27
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH5881 + 344.00 534.32 1.8e-27
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH5881 + 269.50 417.87 9.5e-16
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH5881 + 262.50 395.65 6.4e-15
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH5881 + 262.50 395.11 6.6e-15
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 262.50 395.11 6.6e-15
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 252.50 394.94 6.8e-15
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 255.50 388.58 2.4e-14
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 245.50 372.24 1.9e-13
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 243.00 355.63 4.7e-13
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 243.00 352.35 4.1e-13
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 239.50 347.76 1.2e-12
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 229.00 357.14 3.9e-12
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 228.00 344.98 7.3e-12
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 227.50 329.28 1.4e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:AAH41018 + 227.50 328.46 1.5e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 227.50 328.46 1.5e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 227.50 328.46 1.5e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 224.00 328.46 1.5e-11
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 217.00 313.89 1.2e-10
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 217.00 313.89 1.2e-10
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 216.00 309.64 1.7e-10
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:AAH41018 + 216.00 309.64 1.7e-10
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 199.00 285.77 5.3e-09
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 198.00 284.34 6.5e-09
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 197.50 283.29 7.3e-09
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 197.50 280.05 8.3e-09
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH41018 + 196.50 281.86 9.0e-09
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAH41018 + 196.50 274.74 1.2e-08

```

```

/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH27553 + 195.50 280.24 1.1e
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH27554 + 194.00 278.37 1.4e
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH27549 + 194.00 277.76 1.5e
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH27552 + 193.00 275.90 1.9e
/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:AAH27553 + 191.50 279.91 2.0e

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAO13504

seq_documentation_block:

```

ID      AAO13504 standard; Protein; 667 AA.
XX
AC      AAO13504;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human polypeptide SEQ ID NO 27396.
XX
KW      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW      vaccine; peptide therapy; stem cell; growth factor; haematopoiesis;
KW      tissue growth factor; immunomodulatory; cancer; leukaemia;
KW      nervous system disorders; arthritis; inflammation.
XX
OS      Homo sapiens.
XX
PN      WO200164835-A2.
XX
PD      07-SEP-2001.
XX
PF      26-FEB-2001; 2001WO-US04927.
XX
PR      28-FEB-2000; 2000US-0515126.
XX
PT      18-MAY-2000; 2000US-0577409.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Tang YT, Liu C, Dmanac RT;
XX
DR      WPI; 2001-514838/56.
XX
N-PSDB; AA193435.
XX
PT      Isolated nucleic acids and polypeptides, useful for preventing
PT      diagnosing and treating e.g. leukaemia, inflammation and immune
PT      disorders -
XX
PS      Claim 20; SEQ ID NO 27396; 1399pp + Sequence Listing; English.
XX
CC      The invention relates to human polynucleotides (AA19941-AA19841) and
CC      the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC      cytokine, cell proliferation or cell differentiation or which may induce
CC      production of other cytokines in other cell populations. The
CC      polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC      peptide therapy. The polypeptides have various cytokine-like activities,
CC      e.g. stem cell growth factor activity, haematopoiesis regulating
CC      activity, tissue growth factor activity, immunomodulatory activity and
CC      activity/inhibin activity and may be useful in the diagnosis and/or
CC      treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC      inflammation.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 667 AA;

```

alignment_scores:

Quality: 3298.00 Length: 650
Ratio: 5.113 Gaps: 0
Percent Similarity: 99.231 Percent Identity: 99.077

alignment_block:

US-09-687-230-1 x AAO13504 ..
Align seg 1/1 to: AAO13504 from: 1 to: 667

168 AAGAGCACAAGACAGACAGTCCGACAAACCTTACAGGAGTATGT 217
18 LysLysHisLysLysHisLysSerAspLysHisLeuTyrGlnGlnTyrVal 34
218 AAGAAAGCCCTTGAAGCTGTCTCTCAAGTAGAGGAGGAAAGAGTACCG 267
34 LglLysProLeuLysLeuValLeuLysValLglLysAsnGlnValThrG 51
268 AACTCTCAGCGGACAGTCCGGGACAGCTCCAGCTCTTCCAGACAAA 317
51 LnuSerThrGlySerSerGlyHisAspSerSerLeuPheGlnAspLys 67
318 AACGATCATGACAAACACAAAGACAGAAACGGAAAAAGAAAGAAAG 367
68 AsnAspHisAspLysHisLysAspArgLysArgLysArgLysArgLysG 84
368 AAGAAAGCAGATTCCAGGGGAGAAAGGAGAAAGGAGAAAGGAGATTA 417
84 YglLysGlnLleProGlyGlnGlnLysGlnLysArgLysArgArgValL 101
418 AAGAGATATAAAGAGACGAGATCGAGACCGGTCGAGATGACGACAA 467
101 YsgLnuAspLysLysArgAspArgAspArgValGlnAsnGlnValGln 117
468 AAGATCTCCAGTGTACGCCCTGTGAGATTAGACTTGCCTCTGAGAA 517
118 LysAspLeuGlnCysHisAlaProValArgLeuAspLeuProGlnLys 134
518 GCCTCTCACAAGCTCTTAGCCAAACAGAAAGAGTAGAGACAGACGCC 567
134 sProLeuThrSerSerLeuAlaLysGlnGlnValGlnGlnThrProL 151
568 TTCAGAAAGCTTTGAATCACTGATGAGACAAATTCGAGAGAAAGATCCA 617
151 euGlnGlnLnuLnuAsnGlnLnuMetArgGlnLnuGlnLnuArgLysAspPro 167
618 AGTGGTTTCTTCATTTCTGCTGACTGATTTTATTTGCTCCGCGCTAC 667
168 SerAlaPhePheSerPheProValThrAspPheIleAlaProGlnLysIse 184
668 CATGATCATTAACACCCAAATGATTTTGTACCATGAAAGAAAGATGA 717
184 rMetLleIleLysHisProMetLAspPheSerThrMetLysGlnLysIleu 201
718 AGAACAATGACTATCATGCTCATAGAAAGAACTAAAGATTAACCTCAAACTA 767
201 YSAsnAsnAspTyrGlnSerIleGlnGlnLnuLysAspAsnPhelysLeu 217
768 ATGTGTACTAATGCCATGATTACAAATAAACAGAGACCATTTATTTATA 817
218 MetCysThrAsnAlaMetLleTyrAsnLysProGlnThrIleTyrTyrIly 234
818 AGCTGCAAAAGAGCTGTGCACTCAGAAATGAAATTTCTTACCGAGAAA 867
234 sAlaAlaLysLysLysLeuLnuHisSerGlyMetLysIleLeuSerGlnGln 251
868 GAATTCAGAGCTGAAGCAGACATAGACTTCATGGCTGACTTGACAGAAA 917
251 rGlnLleGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeuGlnLys 267
918 ACTGGAAGCAGAAAGATGGAACACACACCTCAGACAGTGGGGAGAGCG 967
268 ThrArgLysGlnLysAspLysThrAspThrSerGlnSerGlnLysAspL 284
968 AGGCTGTGGCAGAGAGAGAGAGACTCTGAGATGGCCGAAGCAGACAG 1017
284 YglLysTrpGlnArgGlnArgGlnAspSerGlnLysPalagLnuAlaHisA 301
1018 CCTTCAAGAGTCCAGCAAGAAATAAAGAAAGAAAGACAAAGATATGCTT 1067
301 lAphelySerProSerLysGlnAsnLysLysLysAspLysAspMetLnu 317

1068 GAAGATTAAGTTTAAAGCAATAATTTAGAGAGAGACGAGGACGCTTGA 1117
318 GluAspLysPheLysSerAsnAsnLeuGlnArgGlnGlnGlnLnuLnuAs 334
1118 CCGCATCGTAAGAAATCTCGAGAGAAAGCTGACACCGGCTTGTAACA 1167
334 rArgLleValLysGlnSerGlyLysLysLeuThrArgArgLeuValAsnS 351
1168 GTCAGTCCGAATTTGAAGAGAAAGAAACAGATGGAACAACGAGCTTGGA 1217
351 erGlnLysGlnPheGlnArgArgLysProAspLysThrThrThrLnuGly 367
1218 CTTCCTCATCTGTGATCCCATTTAGAGAGACGACGCTACTGCTGGT 1267
368 LeuLnuHisProValAspProLleValGlnGlnProGlnLysCysProVa 384
1268 GAGACGTGGGAATGACAACCTGGAAGACTTGCAGTCGTGGAGTAACCTTGC 1317
384 LArgLeuGlnMetThrThrGlyArgLeuGlnSerGlyValAsnThrLeuG 401
1318 AGGGCTTCAAGAGATTAAGAGACAAAGTCACTCCAGTGTATATTTGG 1367
401 lngLysPheLysGlnAspLysArgAsnLysValThrProValLeuTyrLeu 417
1368 AATTATGGCCCTTACAGTTCTTATGCAACCGCATTAATGACTCCATTTGC 1417
418 AsnTyrGlyProTyrSerSerTyrAlaProHisTyrAspSerThrPheAl 434
1418 AATATCACAAGAGATTTGATTTAATCTAATTCATACCTAAGGGGAAG 1467
434 AsnIleSerLysAspAspSerAspLeuIleTyrSerThrTyrGlnGlnA 451
1468 ACTGTGATCTTCCAAAGTATTCAGCATCCATCAGTTTGTGGCAGCTGC 1517
451 sPserAspLeuProSerAspPheSerIleHisGlnPheLnuAlaThrCys 467
1518 CAAGATTATCCGATATGTCATGGCAGATAGTTTACTGATGTTTAAACA 1567
468 GlnAspTyrProTyrValMetAlaAspSerLeuLnuAspValLeuThrLys 484
1568 AGGAGGGCATTCAGAGACCTCAAGAGATGGAGATGATTCAGTCCCGAAG 1617
484 sGlnGlnLysSerArgThrLeuGlnLnuMetGlnMetSerLeuProGlnA 501
1618 ATGAAGCCCATCTAGACACTTGACACAGAAAGAAATGAGCAGATT 1667
501 sPglLnuGlnHisThrArgThrLeuAspThrAlaLysLnuMetGlnGlnIle 517
1668 ACAGAAATGAGCCACAGGCGCTTGGACTCCAGTACTCAAGACAGGCT 1717
518 ThrGlnValGlnProProGlnArgLeuAspSerSerThrGlnAspArgLe 534
1718 CATAGCGCTGAAGAGCAGTAACAATTTTGGCGTCCAGTTGAAGTTTGG 1767
534 uLleAlaLeuLysAlaValAlaThrAsnPheGlyValProValGlnValPheA 551
1768 ACTGTGAAGAGCTGAATATTTCCAGAGAAACTTGATGAGACCAACAGA 1817
551 sPserGlnGlnLnuLaglnLlePheGlnLysLysLeuAspLysThrThrArg 567
1818 TTGCTCAGGGAATCCAGAGACCCAGAAATGAACGTTTATAGACCAAGACC 1867
568 LeuLeuArgGlnLeuGlnGlnLnuLaglnAsnGlnLnuArgLeuSerThrArgPr 584
1868 CCTGGGACATGATCTGTCTTGGGTCCCGCATC .AGAGAAATGCATC 1916
584 oProProAsnMetIleCysLeuLnuGlnProSerTyrArgGlnMetHisL 601
1917 TTGCTGAACAAGTGAACCAATAATCTTAAAGAAAT .GCACAGCAAGTAACT 1965
601 euAlagLnuGlnValThrAsnAsnLnuLysGlnLeuAlagLnuGlnValThr 617
1966 CCAAGGTATATCGTAAGCACGTAATGAGATTGAAAAAGCAATGGGATTTTC 2015

|||||
618 ProGlyAspIleValSerThrIleGlyValArgLysAlaMetGlyIleSe 634
2016 CATTCTTCCCTCCCTCATGAGAAACACTTTGTGATTGACAGAGACA 2065
634 rIleProSerProValMetGluAsnAsnPheValAspLeuThrGluAspT 651
2066 CTGAAAGAACTTAAAGACGATGTTGATGACTGTCGACCTGGGANGT 2115
651 htrGluGluProLysIlysthrAspValAlaGluCySGlyProGlyGlySer 667
seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAU16626
seq_documentation_block:
ID AAU16626 standard; Protein; 718 AA.
XX
AC AAU16626;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein. Seq ID 1579.
XX
DE Human novel secreted protein. Seq ID 1579.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN W020015322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01341.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

1018 CCTTCAGAGTCCACGAAAGAAATTAAGACAGATATCTT 1067
 1068 GAAGATTAAGTTAAAGCAATATTTAGAGAGAGACGAGACCTTGA 1117
 370 GUAAPLYSPHELYSERASNANLEUGLUAIGLUGLNLNLS 386
 1118 CCGCATCTGGAAGATCTGGAGAAAGTCAAGCGGCTTGTGAACA 1167
 386 PARGLIEVALYSGLUSERGLYLYSLEUTHRAARGLEUVALANS 403
 1168 GTCACTGGGAATTTGAAGAAAGAACAGATGGAACACGCTTGGGA 1217
 403 ERLNCYGLUPHEGLUARGLYSPROAPGLYTHRTHTHTLEUDLY 419
 1218 CTTCCTCATCTGTGATCCCATTTAGAGAGACGAGCTCTGCTGT 1267
 420 LEUENHISPROVALASPROLIEVALIGLUPROGLYTYRCYSPROVA 436
 1268 GAGACTGGGAATGACACTGGAGAGCTTCACTGGAGTGAATCTTGC 1317
 436 IARGLEUGLYMETHTHTGLYARGLEUGLINSERGLYVALANSHTLEUG 453
 1318 AGGGGTCAAGAGATTAAGAAAGCAAGTCACTCCAGTGTATATTTC 1367
 453 IINGLYPHELYSGLUAPLYSARGANLYSVALTHPROVALLEUDLYLEU 469
 1368 AATTATGGCCCTACAGTCTTATGACACCGCATTTATGACTCCATTTGC 1417
 470 ASNTYRGLYPROLYSERSERTYRALAHPROHISTYRASPSERTHPHAL 486
 1418 AATATTCGCAAGATGATTTGATTTAATCTATTTCAACATGAGGAG 1467
 486 AASNLSERTYRASPSERASPLEUULETYSERTHTYRGLYGUA 503
 1468 ACTGATCTTCCAGATGATTTGACATCCATGATTTTGGCCAGCTGC 1517
 503 SPSEASPLEUPROSERASPHESERTIEHLSGLUPHELEUALATHCYA 519
 1518 CAAGATTATCCGTATGTCATGGCAGATAGTTACTGATGATTTTAAACAA 1567
 520 GLNASPYRPROLYVALMETALASPERLEUENUSPVALLEUTHLY 536
 1568 AGGAGGGATTCAGAGACCTTACAGAGATGAGATGTCATGCTGTAAG 1617
 536 SGLYGLYHISSEARTGTHLEUGLUMETGLUMETSERLEUPROGLUA 553
 1618 ATGAAGGCACTAGACACTTGACACAGGAAAGAAATGAGACAGATT 1667
 553 SPGLUGLYHISTHARGTHLEUSPHTHALALYSGLUMETGLU...11E 568
 1668 ACAGAGTAGAGCCACGAGGCGCTTTGACTCCAGTACTCAAGACAGCT 1717
 569 THRGLUVALIGLUPROGLYARGLEUASPERSERTHTGINSAPARGLE 585
 1718 CATGACCCGTGAAGACGATTAACAAATTTGGCGTCCAGTTGAGATTG 1767
 585 UTLALALEUYSALVALTHRASNPHEGLYVALPROVALIGLVALPHEA 602
 1768 ACTCTGAAGAGTGAATATTTCCAGAAAGAACTTGATGAGACACAGCA 1817
 602 SPSERGGLUGLUALAGLUILEPHEGLINLYSLEUASPGULTHTHTARG 618
 1818 TTGCTCAGGAACTCAGAGAACCCAGAAATGAACTTTGAGACACAGACC 1867
 619 LEUENURGLULEUGLINGLUALAGLNASGCLUARGLYSEURTHARGPT 635
 1868 CCCGGGAACATGATCTCTCTTGGGCTCCTCATC AGAAGATGATC 1916
 635 OPTROPHASMETILCYSLLEUGLYPROSERTYRARGGLUMETHSL 652

1917 TTGCTGAACAAGTGACCAATATCTTAAAGATT GCACAGCAAGTAACT 1965
 652 euAlaGlInValThrasnAnLeuYsGLUeAlaGlInGLValThr 668
 1966 CCAAGTATATCTGATGACAGCTATGAGTTGGAAGCAATGGGATTC 2015
 669 PROGLIAPLIEVALSERTHRTYRGLYVALARGYLSALAMETGLYLESE 685
 2016 CATTCCTTCCCGCTCATGAGAAACAACTTTGTGATTTGACAGAAACA 2065
 685 rILeProSerProValMeTGLUASnAnSPheValASpLeuThrGLUASPT 702
 2066 CTGAAGAACTTAAAGACGAGATGTTGATGATGTCGACCTGTGGAAGT 2115
 702 hrglgluprollyserthrasyvalalaglucysgllyproglylser 718
 seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW37947
 seq_documentation_block:
 ID AAW37947 standard; Protein; 589 AA.
 AC AAW37947;
 AS
 DT 14-SEP-1998 (first entry)
 DE Phosphatidylinositol-3' kinase associated protein.
 KW Phosphatidylinositol-3' kinase associated protein; PI3K; PIKAP;
 KW human; signal transduction; cell growth; cancer; restenosis;
 KW therapy; diagnosis.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 151..313
 FT /note="Bromodomain"
 FT Region 516..589
 FT /note="PI3K p85 binding region"
 PN W09820126-A1.
 PD 14-MAY-1998.
 PE 01-OCT-1997; 97WO-US15845.
 PR 01-NOV-1996; 96US-0030103.
 PA (ONXX-) ONXX PHARM INC.
 PI Braselmann S;
 PT WPI: 1998-286942/25.
 DR N-PSDB: AAV29267.
 PT New isolated phosphatidylinositol-3'kinase associated protein -
 PT used to develop products for diagnosis and treatment of cell growth
 PT disorders such as restenosis or cancer
 PS
 PS Claim 10; Page 40-41; 52pp; English.
 CC This polypeptide comprises human phosphatidylinositol-3' kinase
 CC (PI3K) associated protein (PIKAP), a protein that binds to the
 CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
 CC which exhibits a bromodomain. Its amino acid sequence was deduced
 CC from a cDNA clone (see AAV29267) obtained from an HeLa library using
 CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
 CC provides vectors containing nucleic acid sequences that encode
 CC PIKAP or its fragments, host cells, methods for the expression of
 CC PIKAP, and methods for using the products for the diagnosis and
 CC treatment of cell growth disorders such as restenosis or cancer.
 CC Also described is an assay for identifying agonists and antagonists
 CC of PI3K regulation. These include mutant PIKAPs that compete with
 CC native PIKAPs for binding to PI3K, antibodies, and nucleotide

CC sequences that can be used to inhibit or enhance PIPAK gene
 CC expression. Transgenic and knock-out animals are also described.
 XX
 50 Sequence 589 AA;

alignment_scores:
 Quality: 3073.00 Length: 589
 Ratio: 5.217 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-687-230-1 x AAW37947 ..

Align seg 1/1 to: AAW37947 from: 1 to: 589

```

162 ATGGGCAAGACACACAAAGACAACTGGACAAACACCTTCAGAGCA 211
|||||
1 MetGlyLysLysHisLysLysHisLysSerAspLysHisLysLeuTyrGlu 17
212 GATGTGAGAGAGCCCTTGAGCTGGCTCCCAAGTAGAGGAGGAACGAG 261
|||||
17 utryValGluLysProLeuLysLeuValLeuLysValGlyLysAsnGlu 34
262 TCACGCAACTCTCCACGGGACGCTCGGGGACGACTCCAGCCTTCGAA 311
|||||
34 alThrCLeuSerThrGlySerGlyHisAspSerLeuPheGlu 50
312 GACAAAGATCATGACAAACACAGACAGAAACCGGAAAGAGAA 361
|||||
51 AspLysAsnAspHisAspLysHisLysAspArgLysArgLysLysArg 67
362 GAAGAGAGAGAGACAGATTCCAGGGGAGAGAAAGGGAGAAACGAGAA 411
|||||
67 slYsGlyGluLysGlnLeuProGlyGluLysGlyArgLysArgArg 84
412 GACTTAAGAGGATAAAGAGACGAGATCGAGACCGGGTGAGATGAG 461
|||||
84 rValLysGluAspLysLysArgAspArgValGluAsnGlu 100
462 GCGAGAAAGATCTCAGTCTCAGCCGCTGTGAGATTAGACTTGCTCTG 511
|||||
101 AlaGluLysAspLeuGlnCysHisAlaProValArgLeuAspLeuPro 117
512 TGAGAGCCTCTCACAAGCTCTTTAGCCAAACAAAGAGAGAGAGAGAG 561
|||||
117 ogLysProLeuThrSerSerLeuAlaLysGlnGluValGluGln 134
562 CACCCCTTCAGAGGTTGAATCACTGATGAGACAAATTCGAGAGAAA 611
|||||
134 hrProLeuGlnGluAlaLeuAsnGlnLeuMetArgGlnLeuGlnArg 150
612 GATCAAGTCTCTTCTTCAATTCCTGTGACTGATTTATTCCTCTG 661
|||||
151 AspProSerAlaPhePheSerPheProValThrAspPheLeuAlaPro 167
662 CTATCCTCAGATCATTAACACCCCAATGCAATTTAGTACATGAAGAAA 711
|||||
167 yTyrSerMetLeuLeuLysHisProMetAspPheSerThrMetLysGlu 184
712 AGATCAAGACATGACTATCACTGATGAGACAAATTCGAGAGAACTTC 761
|||||
184 yslLeuLysAsnAsnAspTyrGlnSerLeuGlnLeuLysAspAsnPhe 200
762 AAATTAATGTGTAATGCAATGATTTACATTAACACAGAGACATTTA 811
|||||
201 LysLeuMetCysThrAsnAlaMetIleTyrAsnLysProGluThrIleT 217
812 TTATTAAGCTGCAAGAGCTGTTGCACTCAGAGATGAAAAATTCCTGAG 861
|||||
217 rTyrLysAlaAlaLysLysLeuHisSerGlyMetLysIleLeuSerG 234
862 AGAAGAGATTCAGAGCCTGAGACAGAGCATAGACTTCATGGCTGACTG 911

```

```

234 InGUArgLLeuGlnSerLeuLysGlnSerIleAspPheMetAlaAspLeu 250
|||||
912 CAGAAAGCTCGAAAGCAGAAAGATGAAACAGACACTCAGACAGTGGGGA 961
|||||
251 GlnLysThrArgLysGlnLysAspLysThrAspThrSerGlnSerGlyG 267
962 GCAGGAGGCGCTGTGCGAGAGAGAGAGAGAGACTCTGGAGATGCCGAG 1011
|||||
267 uAspGlyGlyCysTrpGlnArgGluArgGluAspSerGlyAspAlaGlu 284
1012 CACAGCCTTCAAGAGTCCACGAAAGAAATTAAGAAAGAAACAAAGAT 1061
|||||
284 lAnHisAlaPheLysSerProSerLysGluAsnLysLysLysAsp 300
1062 ATGCTTGAGATAGTTTAAAGCAATTAATTAGAGAGAGACAGAGAGA 1111
|||||
301 MetLeuGluAspLysPheLysSerAsnAsnLeuGluArgGluGlnGlu 317
1112 GCTTGACCGCATCGTGAAGAAATCTGAGAGAAAGCTGACGAGCGGCTTG 1161
|||||
317 nLeuAspArgIleValLysGluSerClyGlyLysLeuThrArgArgLeu 334
1162 TGAACAGTACGTGCGAATTTGAAGAAAGAAACCGATGGAACAGCAG 1211
|||||
334 aLAsnSerGlnCysGluPheGluArgArgLysProAspLysThrThr 350
1212 TTGGCACTTCTCCATCCTGTGATCCCATTTGAGAGAGCCAGGCTACTG 1261
|||||
351 LeuGlyLeuLeuHisProValAspProIleValGluProGlyTyrCys 367
1262 CCGTGAGACTGGGAAATGACAACTGGAAGACTTCAGTCTGAGAGTAAT 1311
|||||
367 sLeuValArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsp 384
1312 CTTTGAGGGGTTCAAGAGATTAAGAAAGACAAAGTCACTCAAGTGA 1361
|||||
384 hrLeuGlnLysPheLysGluAspLysArgAsnLysValThrProValLeu 400
1362 TATTTGAATTAAGGCCCTPACAGTCTTATGCAACCGCATTAATGACTC 1411
|||||
401 TyrLeuAsnTyrClyProTyrSerSerTyrAlaProHisLysAspSer 417
1412 ATTTGCAATATTCAGAAAGATCATCTGATTTAATCTTCAACCTAG 1461
|||||
417 rPheAlaAsnIleSerLysAspAspSerAspLeuIleTyrSerThrTyr 434
1462 GGGAGACTCTGATCTTCCAAAGTGAATTCAGATCCATGAGTTTGGCC 1511
|||||
434 LysLysPheSerAspLeuProSerAspPheSerIleHisGluPheLeuAla 450
1512 AGCTGCCAAGATTATCCGTAATGTCATGCGACAGATTTTACTGATGTT 1561
|||||
451 ThrCysGlnAspTyrProTyrAlaMetAlaAspSerLeuLeuAspVal 467
1562 AACAAAGAGGCGATTCAGAGCCCTTACAAAGATGAGATGCTCATTCG 1611
|||||
467 uThrLysGlyGlyHisSerArgThrLeuGlnGlnMetGluMetSerLeu 484
1612 CTGAGATGAAGGCCATACTAGACACTTGACACAGAGAAAGAAATGAG 1661
|||||
484 roGluAspGluGlnHisThrArgThrLeuAspThrGlyLysGluMetGlu 500
1662 CAGATTACAGAAATAGACACACAGGCGTTTGAGACTCCAGTACTGAA 1711
|||||
501 GlnIleThrGluValGluProGlyArgLeuAspSerSerThrGlnAs 517
1712 CAGGTCATAGCGCTGAAAGCAGTAACAAATTTGGGCTCCAGTGAAG 1761
|||||
517 pArgLeuIleAlaLeuLysAlaValThrAsnPheGlyValProValGlu 534
1762 TTTTGACTCTGAAGAGCTGAATATTCAGAGAAAGAACTTGATGAGAC 1811
|||||

```

534 a1PheAspSerGluGluAlaGluIlePheGlnLysLysLeuAspGluThr 550
 1812 ACCAGATTCTCAGCAGACTCCAGAGAGCCAGATGACGTTTGACGAC 1861
 551 ThrArgLeuLeuArgGlnLeuGlnGluAlaGlnAsnGlnLysArgSerTh 567
 1862 CAGACCCCTGGGACATGATCTGTCTTGGTCCCTCATCAGAGAAAT 1911
 567 rArgProGlnGlyAsnMetIleCysLeuLeuGlyProSerSerGluLysC 584
 1912 GCATCTTGGTGACAG 1928
 584 yb1LeuLeuAsnLys 589

seq_name: /SID51/gcdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AAW93712

seq_documentation_block:
 ID AAW93712 standard: Protein: 405 AA.

AC AAW93712:
 DT 06-NOV-2001 (first entry)
 DE Human polypeptide, SEQ ID NO: 3652.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping.
 OS Homo sapiens.
 PN EP1130094-A2.
 PD 05-SEP-2001.
 PF 07-JUL-2000; 2000EP-0114089.
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 PA (HELI-) HELIX RES INST.
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI: 2001-524255/58.
 DR N-PSDB: AAK94661.
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 use in genetic manipulation -
 XX Claim 8; SEQ ID NO 3652; 1380pp + sequence listing; English.
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SO Sequence 405 AA;

alignment_scores:
 Quality: 2013.50 Length: 406
 Ratio: 5.034 Gaps: 1
 Percent Similarity: 98.522 Percent Identity: 98.276

alignment_block:
 US-09-687-230-1 x AAW93712 ..
 Align seq 1/1 to: AAW93712 from: 1 to: 405

900 ATGGCTGACCTTGCAGAAAGCTCCGAAAGCAAGAAAGATGCAAGACACCTC 949
 1 MetAlaSerPheGlnGlnLysThrArgLysGlnLysAspLysThrAspThrSe 17
 950 ACAGATGGGAGAGAGAGAGAGCTGCTGGCAGAGAGAGAGAGAGAGAGCTG 999
 17 rGlnSerGlyGlnAspGlyLysTrpGlnArgGlnArgGlnAspSerG 34
 1000 GAGATGCCGAAAGCAGACAGCTTCAAGAGTCCACAGAAAGAAATTAAG 1049
 34 LysArgAlaGlnAlaHisAlaPheLysSerProSerLysGlnAsnLysLys 50
 1050 AAAGCAAGATATGCTGAAGATTAAGTTAAAGCAATATTAAGAGAG 1099
 51 LysAspLysAspMetLeuGlnAspLysPheLysSerAsnAsnLeuGlnL 67
 1100 AGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
 67 gGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 84
 1150 CCAGCGGCTTGTGACAGTCAATGCGAATTTGAAAGAGAAAGACAGAT 1199
 84 hrArgArgLeuValAlaSerGlnCysGlnPheGlnArgGlnArgLysProAsp 100
 1200 GGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1249
 101 GlyThrThrThrLeuGlnLeuGlnHisProValAspProIleValGly 117
 1250 GCCAGGCTACTGCTGTGAGACTGGGAATGACAACTGGAAGACTTCAGT 1299
 117 uProGlyLysCysProValArgLeuGlnMetThrThrGlnArgLeuGln 134
 1300 CTGAGAGTAATCTTTGAGAGGCTTCAAGAGAGATTAAGAGAAAGAAATC 1349
 134 erGlyValAsnThrLeuGlnGlnLysPheLysGlnAspLysArgAsnLysVal 150
 1350 ACTCCAGTATATATTTGAATTAAGGAGGCTTCAAGATTTTGACACGCA 1399
 151 ThrProValLeuTyrLeuAsnTyrGlyProTyrSerSerTyrAlaProH 167
 1400 TTATGACTCCACATTTGCAAAATATCAGCAAGAGATTTGATTTAATCT 1449
 1450 ATTCAAGCTATGGGAGAGACTGTATCTTCCAGAGATTTGACATTCAT 1499
 167 styAspSerThrPheAlaAsnIleSerLysAspAspSerAspLeuIleT 184
 184 ySerThrTyrGlyGlnAspSerAspLeuProSerAspPheSerIleHis 200
 1500 GAGTTTGGGACAGTGGCAAGATTAATCCGATGATGACAGGAGATTAAT 1549
 201 GluPheLeuAlaThrCysGlnAspTyrProTyrValMetAlaAspSerLe 217
 1550 ACTGAGATTTTAAACAAAGAGAGGAGATTCAGAGACCTTACAGAGATG 1599
 217 uLeuAspValLeuThrLysGlnLysHisSerArgThrLeuGlnGlnMetG 234
 1600 AGATGTCTTGGCTGAAGATGAAGCCATTAAGAGACTTGACACAGAGA 1649
 234 lMetSerLeuProGlnAspGlnGlnLysHisThrArgThrLeuAspThrAla 250
 1650 AAAGAAATGAGAGAGATTAAGAAATGAGAGAGAGAGAGAGAGAGAGAG 1699
 251 LysGlnMetGlu...IleThrGlnValGlnProProGlnArgLeuAspSe 266
 1700 CAGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1749
 266 rSerThrGlnAspArgLeuIleAlaLeuLysAlaValThrAsnPheGly 283

```

1750 TTCCAGTTGTAAGTTTGTGACTGTGAAGAGCTGAAATATATCCAGAAGAA 1799
|||||
293 aPrivaGIuValPheaspSerGIuValAGluIlePheGlnIlyS 299
|||||
1800 CTGTGAGAGACCAACGATTGCTCAGAGGAACCTCAGAGAGCCCAAGATGA 1849
|||||
300 LeuAspGIuThrThrArgIleuLeuArgGIuLeuGlnIAGlnInsncI 316
|||||
1850 ACGTTTGAGACCAACGACCCCGTGGGAACATGATCTGTCTTGGGTCCCT 1899
|||||
316 uArgIleuSerThrArgProProProAsnMetIleCysIleuLeuGIyProS 333
|||||
1900 CATC AGAGAAATGCATCTGTGTGAACAGTACCAATTAATCTTAAGAA 1948
|||||
333 eTtYtArgIuMetHsIleuAlaGIuGlnValThrAsnInsleuIySgIu 349
|||||
1949 TTGCACAGCAAGTAACCTCCAGTGATATCTAAGCACCTATGAGAGTTG 1997
|||||
350 LeuAlaGIuGlnIValThrProGIyAspIleValSerThrTtYGIyAlaR 366
|||||
1998 AAAAGCAATGGGATTTCCATTCCCTCCCGTCATGAAAAACATTGG 2047
|||||
366 gIySaIaMetGIyIleSerIleProSerProValMetGIuAsnAspPheY 383
|||||
2048 TGGATTGTGACAGACACTGAGAAACCTAAAGACGATGTGATGNG 2097
|||||
383 aIAspIleuThrGIuAspThrGIuIleProIySfThraspAlaIAGlu 399
|||||
2098 TGTGCACCTGTGTGAAGT 2115
|||
400 CysGIyProGIyGIySer 405
|||||
seq_name: /SIDS1/gcdata/hold-geneseq/gene-seq-emb1/AA2001.DAT: AAB93765
seq_documentation_block:
ID AAB93765 standard; Protein: 351 AA.
XX
XX AAB93765;
XX
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:13461.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI, 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13461; 2537bp + CD ROM; English.
XX

```

```

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 351 AA:

```

```

alignment_scores:
Quality: 1716.50      length: 352
Ratio: 4.975          gaps: 1
Percent Similarity: 98.011 Percent Identity: 97.727

```

alignment_block:

US-09-687-230-1 x AAB93765 ..

Align seg 1/1 to: AAB93765 from: 1 to: 351

```

1062 ATGCTTGAAGATTAAGTTTAAAGCAATATTTAGACAGACAGAGAGCA 1111
|||||
1 MetLeuGIuAspIySphelySerAsnAsnLeuGIuArgGIuInsncI 17
|||||
1112 GCTTGACCCATCGTGAAGAAATCTGAGAGAAAGCTGACAGAGCGGCTTG 1161
|||||
17 nIeuAspArgIleValIySgIuSerGIyIySleuThrArgArgIeU 34
|||||
1162 TGAACAGTCAGTCGCAATTTGAAAGAAAGAAACACAGATGCAACACAGC 1211
|||||
34 aIAsnSerGIuInsncIuPheGIuArgArgIySphArgIyThrThr 50
|||||
1212 TTGGGACTTCTGCATTCCTGTGATCCCATTTGAGAGAGACCGAGCTACG 1261
|||||
51 LeuGIyLeuIleuHsIProValAspProIleValGIyGIuProGIyTtY 67
|||||
1262 CCTGTGAGACTGGCAATGACAACCTGAGACACTTCAGTGTGAGTGAATA 1311
|||||
67 sProValArgIleuGIyMetThrThrGIyArgIeUInsncIyAlaIns 84
|||||
1312 CTTTGCAAGGGGTTCAAGAGATTAAGAGAACAACTCACTCCAGTGTGA 1361
|||||
84 hIleuGIuInsncIyPheIySgIuAspIyArgAsnIyValThrProVal 100
|||||
1362 TATTGGAATTATGGGCCCTACAGTTCTTATGACACCGCATTAAGATCCAC 1411
|||||
101 tYrIleuAsnTtYGIyProIySerSerIyAlaProHsIstYrAspSer 117
|||||
1412 ATTTCGAATATACCAAGAGATGATTTGATTAATCTTAACCTTANG 1461
|||||
117 rHehIaAsnIleSerIyAspAspSerIleuIeTtYrSerThrTtYc 134
|||||
1462 GGAAGACTGTGATCTTCCAGATGATTTTCAGATCAGATGAGTTTGGCC 1511
|||||
134 tYGIuAspSerAspIeuProSerAspPheSerIleHsIcIuPheIeua 150
|||||
1512 ACGTGCAAGATTATCCGATTCATGCGAGATAGTTTACTGGAGTGT 1561

```



```

305 CTTGAGAGAAAAACGATCAT.....GACAAACAGAGAGAGAAAGC 348
      ::::::::::::::::::::
59 rTyAspAspArgSerAspHisGluArgGluArgHisLysGluLysLysL 76
349 GGAAGAGAGAGAGAGAGAGAGAGAG.....ATTCCAGGGGAGAGAA 392
      ::::::::::::::::::::
76 yAluLysLysLysLysSerGluLysGluLysHisLysLeuAspArgGlu 92
393 AAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 442
      ::::::::::::::::::::
93 .....ArgArgLysArgLysGluLysLysLysArgLysArgGluArg 106
443 AACACGGGGGAGAGAGATGAGGAGAGAA...AAAGATCTCCAGTGCACGCC 489
      ::::::::::::::::::::
106 gEluHisCysAspThrGluGluGluLysLysAspAspProGluLysL 123
490 CTGTGAGATTAGACTTCCCTCCGAGAGAGAGAGAGAGAGAGAGAGAG 539
      ::::::::::::::::::::
123 yAlaGluValGluProProProAspArgProValArgAlaCysArgThr 139
540 AAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 589
      ::::::::::::::::::::
140 GluProAlaGluLysGluLysSerThrProIleGluGluLysLysHis 156
590 GATGAGACAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 639
      ::::::::::::::::::::
156 eLeuArgGluLysGluLysArgLysAspProHisGlyPhePheAlaPhePro 173
640 TGACTGATTTATTTGCTCTGCTGCTACTGATGATGATGATGATGATG 689
      ::::::::::::::::::::
173 aLThrAspAlaIleLeuArgProGlyTyrSerMetIleIleLysHisPro 189
690 GATTTTATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739
      ::::::::::::::::::::
190 AspPheGlyThrMetLysAspLysIleValAlaAsnGluTyrLysSer 206
740 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
      ::::::::::::::::::::
206 lThrGluLysLysAlaAspPheLysLeuMetCysAspAsnAlaMetThr 223
790 ACAATTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
      ::::::::::::::::::::
223 yTAsnArgProAspThrValTyrTyrLysLeuAlaLysLysIleLeuHis 239
840 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872
      ::::::::::::::::::::
240 AlaGlyPheLysMetSerLysGluAlaAlaLeuLeuGlyAsnGluAs 256
873 .CAGAGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
      ::::::::::::::::::::
256 pThrAlaValIleGluLysProValIleProValAlaIleProValGlu 270
922 GAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 971
      ::::::::::::::::::::
270 ..... 270
972 TGCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
      ::::::::::::::::::::
270 ..... 270
1022 CAAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
      ::::::::::::::::::::
271 .ValGluThrAlaLysLysSerLysLysProSerArgGluValIleSer 287
1072 ATTAGCTTTAA.....AGCAATATAT 1091
      ::::::::::::::::::::
287 ySMeTPheGluProGluLysAsnAlaCysSerLeuThrAspSerThrAla 303
1092 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
      ::::::::::::::::::::
304 GluGluHisValIleLeuAlaLeuValGluHisAlaAlaAspGluAla 320
1142 AAAGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1191

```

```

      ::::::::::::::::::::
320 pArgIleAsnArgPheLeuProGluGlyLysMetGlyTyrLeuLysArg 337
1192 AACCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
      ::::::::::::::::::::
337 snGlyAspGlySerLeuLeuTyrSerValIleAsnThrAlaGluProAsp 353
1242 GTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1291
      ::::::::::::::::::::
354 AlaAspGluGluGluThrHisProValAspLeuSerSerLeuSerSer 370
1292 ACTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1341
      ::::::::::::::::::::
370 sLeuLeuProGlyPheThrThrLeu...GlyPheLysAspGluArgArg 386
1342 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1391
      ::::::::::::::::::::
386 snLysValThr.....PheLeuSer...SerAlaThrThrAlaLeu 398
1392 GCACCGCATTTATGACTCCACATTTGCAAAATATCAGCAAGAGATGAT 1441
      ::::::::::::::::::::
399 SerMetGluAsnAsnSerValPheGlyAspLeuLysSerAspGluMetG 415
1442 TTTATCTATTTCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1491
      ::::::::::::::::::::
415 uLeuLeuTyrSerAlaTyrGlyLysArgLysGluThrGlyValGlnCysAlaLeu 432
1492 GCATCCATGATGATTTTGGCCAGCGCAAGATTTATCCGCTATGCAAGCA 1541
      ::::::::::::::::::::
432 eLeuGluGluLysPheValLysAspAlaGlySerTyrSerLysValVal 448
1542 GATAGTTTACTGAGATGTTTAAACAGAGAGAGAGAGAGAGAGAGAGAG 1591
      ::::::::::::::::::::
449 AspAspLeuAspGlnIleThrGlyLysPheIleSerArgThrLeuPhe 465
1592 AGAGATGAGAG.....ATGTCATTCCTCGAAGATGAGAG 1623
      ::::::::::::::::::::
465 eGluLeuLysGluArgArgAsnValProMetLysProProAspGluAla 482
1624 GCCATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1670
      ::::::::::::::::::::
482 yAlaGlyAspThrLeuGlyAspSerSerSerSerValLeuGluPheMet 498
1671 GAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720
      ::::::::::::::::::::
499 SerMetLys.....SerTyrProAspValSerVal 508
1721 AGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770
      ::::::::::::::::::::
508 lAspIleSerMetLeuSerSerLeuGlyLysValLysLysGluLeuAsp 525
1771 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1820
      ::::::::::::::::::::
525 roAspAspSerHisLeu.....AsnLeuAspGluThrThrLysLeu 538
1821 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1870
      ::::::::::::::::::::
539 LeuGluAspLeuHisGluAlaGlnAlaGluArgGlyLysArgProse 555
1871 TGGAGACATGATCTGCTGTGGCTCCCTCA 1901
      ::::::::::::::::::::
555 rSerAsnLeuSerSerLeuSerAsnAlaSer 565
seq_name: /SID1/gcgdata/hold-geneseq/genesep-emb1/AA2001.DAT:AA38835
seq_documentation_block:
ID AA38835 standard; Protein: 597 AA.
xx
xx AA38835;
xx
xx AC
xx AC
xx 22-Oct-2001 (first entry)
xx
xx Human polypeptide SEQ ID NO 1980.

```


[illegible]

seq_documentation_block:	
ID	AU016213 standard; Protein: 199 AA.
XX	
AC	AU016213;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human novel secreted protein, Seq ID 1166.
XX	
KM	Human; immunosuppressive; antiarthritic; antirheumatic;
KM	cytostatic; cardian; vasotropic; cerebroprotective; nootropic;
KM	neuroprotective; antibacterial; virucide; fungicide; optalmalogical;
KM	vulnery; secreted protein; rheumatoid arthritis;
KM	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KM	cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KM	nervous system disorder; Alzheimer's disease; infection; ocular disorder
KM	corneal infection; wound healing; epithelial cell proliferation;
KM	skin ageing; food additive; preservative; antiproliferative.
XX	
OS	Homo sapiens.
XX	
PN	WO200155322-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-0501341.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0198874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226688.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 25-SEP-2000; 2000US-0234999.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239393.
 PR 13-OCT-2000; 2000US-0239397.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251859.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-488783/53.

N-PSDB; AAS26200.

PT New nucleic acid molecules encoding 461 human secreted proteins for
 diagnosing, preventing, treating or ameliorating medical conditions and
 used as food additives or preservatives -

Claim 11; SEQ ID NO 1166; 980pp; English.

CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

alignment_scores:

Quality:	861.50	Length:	489
Ratio:	4.351	Gaps:	4
Percent Similarity:	40.491	Percent Identity:	40.082

alignment_block:

US-09-687-230-1 x AAU16213 ..

Align seg 1/1 to: AAU16213 from: 1 to: 199

210 GAGTATGTAGAGAGCCCTTGAAGCTGTCTCAAGTAGAGGAGGAGCA 259
 ||||||||||||||||||||||||||||||||||||||||||||||||
 1 GlturyValGltulysProleuNylsleuValGlyGlyAsnI 17
 260 AGTCACCGAACTCTCCACGGGCAAGCTCGGGGACGACTCAGCCTTTCG 309
 ||||||||||||||||||||||||||||||||||||||||||||||||
 17 uValThrGltuleuSerThrGlySerSerGlyHisPaspSerSerIleupheG 34

```

310 AAGACAAAACGATCATGACAAACACAGACAGAAACGGAAAAAGAGA 359
|||||
34 lnsprlysnspnhsprlyshlsysasprlylsarplylsarplysarg 50
|||||
360 AAGAAAGGAGAGAGATTCAGAGGGGAAAGAAAGGAGAAAGGAG 409
|||||
51 LysLysGlyGlyLysGlnLeuProGlyGlnGlyLysGlyLysArg 67
|||||
410 AAGAGTTAAGGAGATAAAAAGAGGAGATCGAGACCGGTGAGAAATG 459
|||||
67 ghrGValLys..... 70
460 AGGCAAAAAAGATCTCCAGGTGACAGCCCTGTGAGATTAGACTTGCT 509
|||||
70 ..... 70
510 CCTGAGAGCCTCTCACAAGCTCTTAGCCAAACAGAAAGATAGACA 559
|||||
70 ..... 70
560 GACACCCCTTCAAGAGCTTTGAATCACTGATGACACAATTGACAGAA 609
|||||
70 ..... 70
610 AAGATCCAGTGTCTTTCTTTTCATTTCCGTGACGTATTATTGCTCCT 659
|||||
70 ..... 70
660 GGCTACTCATGATCATTTAAACCCCAATGATTTAGTACCATGAAAGA 709
|||||
70 ..... 70
710 AAAGATCAAGAACATGACTATCATCTCCATAGAGAACTAAAGATTA 759
|||||
70 ..... 70
760 TCAACTAATGTGTAATAATGCCATGATTTACAATAAACCCAGACCA 809
|||||
70 ..... 70
810 TATATTAAGCTGCAAAAGAGCTGTGCACTCAGAAATGAAATTTCTAG 859
|||||
70 ..... 70
860 CCAGGAAGAAATTCAGAGCCTGAGACAGACATAGACTTCATGCTGACT 909
|||||
70 ..... 70
910 TGCAGAAAACTCGAAAAGCAGAAAGATGAGACAGACCTCACAGAGTGG 959
|||||
70 ..... 70
960 GAGACGAGAGCTGCTGCGCAGAGAGAGAGAGACTCTGAGATGCCGA 1009
|||||
70 ..... 70
1010 AGCACAGCCTTCAAGAGTCCACAGAAAGAAATTAAGAAAGAACAAAG 1059
|||||
70 ..... 70
1060 ATATGCTTGAAGATTAAGTTTAAAGCAATATTTAGAGAGAGAGAGAG 1109
|||||
70 ..... 70
1110 CAGCTTGACGCGATCGTGAAGAAATCTGAGAGAAAGCTGACAGGCGCT 1159
|||||
70 ..... 70
1160 TGTGAACAGTCAATTTGAAGAAAGAAACAGATGAGAAACAGCA 1209
|||||
71 .....CysGlnPheGlnArgArgLysProAspLysThrThr 83

```

```

1210 CGTTGGGACTTCTCATCTGTGATCCATTTGTAGAGAGCCAGGCTAC 1259
|||||
83 hrlxuglyleuLeuHlsProValAspProIleValGlyGlnProGlyTyr 99
|||||
1260 TGGCTGGTGAAGTGGGATGACAACTGGAAGACTTCAGTCTGAGTGA 1309
|||||
100 CysProValArgLysGlnMetThrThrGlyArgLysGlnSerGlyValAs 116
|||||
1310 TACTTGCAGCGGCTCAAGAGATTAAGAGAAAGCAAGTCCAGCTG 1359
|||||
116 nhrleuGlnGlyPheLysGlnAspLysArgAsnLysValThrProVal 133
|||||
1360 TATATTTGAATTAATGGCCCTACAGTCTTATGACAGCATTAATGAT 1409
|||||
133 eutyrlxusnryrGlyProLysSerSerTyrLalaProHsTyrAspSer 149
|||||
1410 ACATTTGCAAAATATCAAGAGATGATTTGATTAATCTAATCAACCA 1459
|||||
150 ThrPheLalaAsnLleSerLysAspAspSerAspLeuIleTyrSerThr 166
|||||
1460 TGGGGAAGACTCGATCTTCCAAAGTATTCAGCATCCATGAGTTTGG 1509
|||||
166 rclGlnAspSerAspLeuProSerAspPhe.Ser..... 177
|||||
1510 CCAGGTGCCAAGATTATCCGATATGTCATGACAGATAGTTACTGAT 1559
|||||
177 ..... 177
1560 TTAACAAAAGAGGGCATTCAGAGCCCTTACAAGAGATGAGATGCA 1609
|||||
178 .....H1SC 179
1610 GCCTGAAGATGAAGCCATACTAGACACTTGACACAGGAAAGAAATG 1659
|||||
179 yslxulysmetLysAlaIleLeuGlyHisLeuThrGlnGlnLysLysTrp 195
|||||
1660 AGCAGATTACAGAG 1674
|||||
196 ...ArgLeuGlnLys 199
seq_name: /SIDSL/gcdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:ABB63028
seq_documentation_block:
ID ABB63028 standard; Protein: 861 AA.
AC ABB63028;
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 15876.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
PE 23-MAR-2001; 2001MO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PA (PEKE ) PE CORP NY.
PI Venter JC, Adams M, Li PMD, Myers EW;
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07131.
XX New isolated nucleic acid detection reagent for detecting 1000 or more

```

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 15876; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SO Sequence 861 AA;

alignment_scores:
 Quality: 738.50 Length: 824
 Ratio: 1.656 Gaps: 24
 Percent Similarity: 54.126 Percent Identity: 28.034

alignment_block:
 US-09-687-230-1 x ABB63028 ..

Align seg 1/1 to: ABB63028 from: 1 to: 861

168 AAGAGCACAAGACACAAGTCGGAC...AAGACCTCTACGAGAGTA 214
 |||||
 5 LysLysHisLysLysAsnLysSerGluArgGluLysTyrGluLys 21
 215 T.....GTAGAGAGCCCTTGAAGCTGTC 240
 |
 21 TserGlnHisGlnAspProAlaGlnLeuGlnArgGlyLeuLysLeuLeu 38
 241 TCMAAGTAGAGGGAAGACAGTCACCGAAGCTCCACGGGAGCTGGGG 290
 |||||
 38 eulysvalglserAsnAlaThrProGluTyrSerAlaAsnSerPromet 54
 291 CACGAC.....TCCAGCCGCTT 307
 ||||
 55 valAspGlyLysProThrAlaAlaGluAlaMetMetSerProValPr 71
 308 CGNAGCACAAGACATGATGACAAACAGACAGAGAAAGCGGAAAAAGA 357
 |||||
 71 ogLugluLeuGlnAspHisGlnGlnHisArgGluArgHisLysLysSer 88
 358 GAAAGAGAGAGAGAAAGATTCAGGGGAGAAAGGGGAGAAAAACGG 407
 :|||
 88 yslLysLysLysLys.....LysAspArgGluLysLysHis 100
 408 AGAAGAGTTAAGAGATATAAAGAACCGAGATCGAGACCGGGTGAGAA 457
 :|||
 101 LysHisHisLysGlu...LysArgHisArgSerArgAspArgHisArgAs 116
 458 TGAGGCGAAG.....AAGATCTCCAGTGC 483
 :|||
 116 palagLysSerAspGlnAspMetMetAlaGlyAlaAspAspAlaAlaCys 133
 484 AC..... 485
 133 erGlyPheAlaProSerSerValAlaProProAlaAlaAspProAspSer 149
 485 485
 150 SerGlnAspGlyPheSerPheMetAspAspAspGlnSerGlnProLeuPr 166
 485 485
 166 ogLysAsnIleLeuPhePheAlaGlyIleThrThrAspAsnSerProSerA 183

486GCCCTGTGAGATTAGAC..... 503
 |||||
 183 snCyProValThrLysProIleAlaProArgLysLeuAspAlaLeu 199
 503 503
 200 MetGlySerSerProAsnSerSerSerLeuGlnSerSerLeuGlyLe 216
 504TTGCCT..... 509
 216 ulleGlySerSerProThrLysProLeuProAspLeuLeuIleProSer 233
 509 509
 233 roSerThrProGlyGlyAlaAsnSerLeuAsnAlaLeuThrProLysAla 249
 510CCTGAGAGCCTTCACAGAGCTTTAGCCAAACAGAA... 548
 |||||
 250 LeuGluAlaProLysThrProSerSerSerSerGlnSerGlnArgGluPr 266
 549GAAGTAGAACAGACACCCCTTCAGAGAG 576
 :|||
 266 oArgSerCyValLeuLysLeuLysGlnGlnLysSerProLeuAsnLysL 283
 577 CTTTGAAATCACTGATGAGACAATTGCAGAGAAAAGATCCAAAGCTTTC 626
 |||||
 283 eulLeuGlnHisLeuLeuArgPheLeuGluLysArgAspProHisGlnPhe 299
 627 TTTTCATTTCCCTGACTGATTTATGCTCCCTGCTGCTGCTGCTGCTGCT 676
 |||||
 300 PheAlaThrProValThrAspAspMetAlaProGlyLysSerIleI 316
 677 TAAACCCCAATGATTTTAGTACCATGAAGAAAGATCAAGAACATG 726
 |||||
 316 eSerArgProMetAspPheSerThrMetArgGlnLysIleAspAspHisG 333
 727 ACTATGCTCCATGAAAGAACTAAAGATTAACCTCAACTATGTGACT 776
 |||||
 333 LysThrAlaLeuThrGluPheThrAspAspPheLysLeuMetCyGlu 349
 777 AATGCCATGTTTACATTAACCAACAGACACCATTTATATTAAGCTGCAA 826
 |||||
 350 AsnAlaIleLysTyrAsnHisValAspThrValTyrAsnLysAlaAlaLys 366
 827 GAAGCTGTCACACTCAGAAATGAAGATTTAGCCAGAAAGAAAT...C 873
 |||||
 366 sArgLeuLeuGlnValGlyMetLysHisLeuGlnProGlnAsnLeuMet 383
 874 AGAGCTGAAGCAGACGATAGACTTCATGCTGCTGCTGCTGCTGCTGCA 923
 :|||
 383 rGSerLeuLysProLeuSerGlyTyrMetArgGluLeu.....Thr 396
 924 AAGCAGAAAGATGGAACAGACACCTCACAGAGTGGGGAGAGAGGCTG 973
 :|||
 397 AlaArgGluLeuGlyPheGluLeuSerSerAsnAsp..... 408
 974 CTGGCAGAGAGAGAGAGAGACTCTGGAGAT.....G 1005
 :|||
 409 MetSerArgGlnAsnAsnAspSerAlaAspGluGlyAlaSerThrGlyA 425
 1006 CCGAAGCAGACAGCCTTCACAGCTCCACAGAAAGAAATTAAGAAAGAC 1055
 |||||
 425 lArgLysLysProArgThrProAlaGlnLeuGlnGlnGlnArgLysArg 441
 1056 AAGATATGCTTGAAGATTAAGTTAAAG.....AA 1087
 |||||
 442 ThrLeuArgLeuGlnAsnAlaProLysThrHisPheGluProTyrValAs 458
 1088 TAAATTAAGAGAGAGAGAG.....GAGCAGCTTGACCGCATCTGAAG 1131
 :|||
 458 pAspLeuThrGlyGlnGluIleLeuAlaGlnValGlnAsnAlaAlaGlnG 475

```

1132 AATCTGAGAGAAAGCTGACACGCGCTTGTAACAGTCACTGCATTT 1181
      ::::::::::::::::::::
475  InAlaLysGlnArGValAsnAlaLysAsnAlaHisLysMetGlyPhe 491
      ::::::::::::::::::::
1182 GAAAGAAAGAAAGACATGATGAAACAGACGTTGGACTCTCCATCCTGT 1231
      ::::::::::::::::::::
492 LeuArGlnMetLysAspGlyThrThrThrLeuAsnLeuValIleLysG 508
      ::::::::::::::::::::
1232 GGATCCCATTTGTAGAGAGCCAGGCTACTGCTGTGAGACTGGAGATGA 1281
      ::::::::::::::::::::
508 uGlu.....AsnGluGlyProGluuArgValValThrIleGlyAspL 522
      ::::::::::::::::::::
1282 CAACGTGAGAGACTTCAGTCTGAGAGTCACTTTCACAGGGCTTCAAGAG 1331
      ::::::::::::::::::::
522 euValGlyLysLeuGlnLysGlySerAlaGlnLeuGlnThrArgGlnVal 538
      ::::::::::::::::::::
1332 GATAAAAGAAAGAAAGTCACTCCAGTGTATATGAAATATGAGCCCTTA 1381
      ::::::::::::::::::::
539 AspLysArGAsnAlaValAlaArgThrValLysSerLeuAsnTyrGlyAla 555
      ::::::::::::::::::::
1382 CAGTCTTATGACCGCATTTATGACTCCACATTTGCCAAATATCAGCAGG 1431
      ::::::::::::::::::::
555 eAlSerPheAlaProThrPheAspSerArgPheSerThrLeuSerAlaG 572
      ::::::::::::::::::::
1432 ATGATTCGATTTATCTATCTCAACCTATGGGAGAGACTCTGATCTTCCA 1481
      ::::::::::::::::::::
572 IuGluThrGlnLeuValAlaLeuArgThrTyrGlyAspAlaSerSerAlaG 588
      ::::::::::::::::::::
1482 AGTGATTTTCAGCATTCATGAGTTTGGCCACGTCGCAAGATATTCGCTA 1531
      ::::::::::::::::::::
589 TyrAlaGlnSerIleLeuGlnPheThrLysAsp...SerAsnTyrGlyTh 604
      ::::::::::::::::::::
1532 TGTCATGGCAATAGTTTACTGGATGTTTAAACAAGAGGGGCACTTCCA 1581
      ::::::::::::::::::::
604 rThrIleAlaSnnglyLeuLeuAspIleLeuThrAsnGlyLysIleSerL 621
      ::::::::::::::::::::
1582 GGACCCCTA.....CAA 1592
      ::::::::::::::::::::
621 ySserLeuAspGlnLeuTyrAsnMetGlnLeuHisSerTyrGlnGlnArg 637
      ::::::::::::::::::::
1593 GAGATGAGATGTCATTCCTGTAAGATGAGAGCCATCTAGACACTTGA 1642
      ::::::::::::::::::::
638 GluIleGlyLysCysPheGlnGlnGlu.....G 648
      ::::::::::::::::::::
1643 CACAGGAAAGAAATGGACGATTTACAGAGTACAGCCACCGAGGCGTT 1692
      ::::::::::::::::::::
648 uThrSerSerGlnGlnGlnThrAlaGlnIleGlnGlnGlyLysGln 665
      ::::::::::::::::::::
1693 TGGACTCCAGTCAAGACAGGCTCATAGGCTGAAGAGTACATCAAT 1742
      ::::::::::::::::::::
665 ySThrLysAsnThrHisValAspPheLysGlyLeuGlnSerLeuGlyGlu 681
      ::::::::::::::::::::
1743 TTGGCGCTTCAGTTGAATTTTGGACTCTGAGAAAGCTGAATA..... 1787
      ::::::::::::::::::::
682 LeuGlyIleAspValSerPheLeuAspGlyMetGluAlaGlnMetLys 698
      ::::::::::::::::::::
1788 .....TTCCAGAGAAAGTGTATGACGCCACCAAT 1818
      ::::::::::::::::::::
698 rTyrGlnLeuAsnArgLysMetHisGlnHisLysSerGlnInsLeuThrL 715
      ::::::::::::::::::::
1819 TGCACAGAGACTCCAGAGACCCAGATGAAGCTTTGAGACACCGACCC 1868
      ::::::::::::::::::::
715 euIleGlyLysLeuArgValAlaGlnHisAspArgLysSer...GlnPro 730
      ::::::::::::::::::::
1869 CCTGGGAAACATGATCTGTCTCTGGGCTCCATC...AGAGAAATGCA 1914
      ::::::::::::::::::::
731 LeuProAsnHisLeuGlyLeuValGlnProAlaGlnGlnGlnGlnGln 747
      ::::::::::::::::::::
1915 TCTGTGTCAGAAAGTGAACAATTAATCTTAAGA ATTGCACACGCAAGTAA 1963
      ::::::::::::::::::::
747 nThrAlaGlnGlnLeuThrGlnGlnIleSerAspLeuValLysLysLeuP 764
      ::::::::::::::::::::
1964 CTCACAGGTGATATGTGTAAGCACGATATGAGAGTTCGAAAAAGCAATGGCGATT 2013

```

```

      ::::::::::::::::::::
764 roProSerAlaIleAlaAspProTyrAlaLeuArgLysAlaMetGlyMet 780
      ::::::::::::::::::::
2014 TCC.....ATTCTTCCCGCCGTCATGAGAAACACTTTGTGATTT 2054
      ::::::::::::::::::::
781 SerTyrAlaGlyLeuProProAlaArgProValSerProArgValGlnLe 797
      ::::::::::::::::::::
2055 GACAGAGACACTGAGAAACCT 2076
      ::::::::::::::::::::
797 uProGlnLeuLeuGlnGlnPro 804
      ::::::::::::::::::::

seq_name: /SIBS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT: AAB95421
seq_documentation_block:
ID AAB95421 standard; Protein: 501 AA.
AC AAB95421;
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:17823.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
PE 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 8; SEQ ID 17823; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

```

CC of the present invention.

XX Sequence 501 AA:

alignment_scores:

Quality: 684.00 Length: 473

Ratio: 2.178 Gaps: 11

Percent Similarity: 66.385 Percent Identity: 34.672

alignment_block:

US-09-687-230-1 x AAB95421 ..

Align seq 1/1 to: AAB95421 from: 1 to: 501

```

549 GAAGTAGAGACAGACACCCCTTCAGAAAGCTTTGATCAACTGATGAGACA 598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 GlnuSnGluSerThrProIleGlnGlnLeuLeuLunHisPheLeuAlrG1 35
599 ATTGCAGAGAAAGATCCAAAGTCTTCTTTTCATTTTCCTGACTGATT 648
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 nleuGlnArgLysAspProHisGlyPhePheAlaPheProValThrAspA 52
649 TTATTCCTCCGCTACTCCATGATCATTAACACCAATGATTTAGT 698
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 lalleAlaProGlyTyrSerMetIleIleLysHisProMetAspPheLys 68
699 ACATGAAAGAAAGATCAAGACATGATCATGCTCATGAGAAAGACT 748
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 ThrMetLysAspLysIleValAlaSnclutryLysSerValThrGluPhe 85
749 AAAGCATTAATCTCAACTATGTGTACTATGCCATGATTTACAAATAAC 798
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85 eLyAlaAspPheLysLeuMetCysAspAspAlaMetThrTyrAsnAspR 102
799 CAGAGACACATTTATTAAGCTGCAAGAGCTGTGCTACTGACCTAGGATG 848
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 roAspThrValTyrTyrLysLeuAlaLysLysIleLeuHisAlaLysPhe 118
849 AAAATTTCTTACCCAGAAAGATTCAGACCTGAGAGACATGAGACTT 898
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 LysMetMetSerLysGluArgLeuLeuAlaLeuLysArgSerMetSerPhe 135
899 CATGGCTGACTTGCAGAAACTCCAAAGCAGAAAGATGGAACAGACACT 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 eMetGlnAspMet.....AspPheS 142
949 CACAGAGGCGGAGAGAGAGCTGCTGCGCAGAGAGAGAGAGAGACT 998
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 eGlnGlnAlaAlaLeuLeuLys.....AsnGlnAspThr 153
999 GGAGATGCCGAGAGACACAGCCTTCAGAGATGCC.....Ag 1033
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 AlaValGluLuproValProGluValAlaProValGluValThrAl 170
1034 CAAGAATAATAAAAAGAAAGACAAAGATGCTTGAAGATTAATTAA. 1082
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 aLysLysSerLysLysProSerArgGluValIleSerCysMetPheGluP 187
1083 .....AGCAATTAATTTAGAGAGAGAG 1103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 roGluGluAsnAlaLysSerLeuThrAspSerThrAlaGluLunHisVal 203
1104 CAGGAGACGCTTGACCCGATCGTGAAGAGATCTGGAGAAAGCTGACGAG 1153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 LeuAlaLeuValGluHisAlaAlaAspGluAlaArgAspArgIleAsnArg 220
1154 GCGGCTTGTGAACAGTCAAGTGGAAATTGAAAGAAAGAAAGACAGATGAA 1203
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 gPheLeuProGluGlyLysMetGlyTyrLeuLysArgAsnGluAspGlyS 237
1204 CAACGACGTTGGAGACTCTCCATCTGTGAGATCCATTTGTAGAGAGCA 1253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

237 erLeuLeuTyrSerValValAsnThrAlaGluProAspAlaAspGlu 253
1254 GGTACTGCTCGGTGAGACTGGGATGACAACTGGAAAGCTTCAGTCTGG 1303
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GluThrHisProValAspLeuSerSerLeuSerLysLeuLeuProG1 270
1304 AGTGAAATTAATTTGCAAGGATCAAGAGATGAAGAAAGAAAGAGCTC 1353
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
270 yPheThrThrLeu...GlyPheLysAspGluArgAsnGlyValThr. 285
1354 CAGTGTATTAATTTGAATTTATGAGCCCTACAGTCTTATGACCCGATTA 1403
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 .....PheLeuSer...SerAlaThrThrAlaLeuSerMetGlnAsn 298
1404 GACTCCACATTTGCAAAATATCAGCAAGAGATGATTCGATTTAATCTATTC 1453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 AsnSerValPheGlyAspLeuLysSerAspLysMetGluLeuLeuTyrSe 315
1454 AACCTATGGGAGAGACCTGATCTTCCAGATGATTTCCAGATGATGAGT 1503
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 lAlaTyrGlyAspGluThrGlyValGlnCysAlaLeuSerLeuGlnGluP 332
1504 TTTTGGCCACGCTGCCAAGATTAATCCGATGTCATGAGCAGATAGTTACTG 1553
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 heValLysAspAlaGlySerTyrSerLysValValAspAspLeuLeu 348
1554 GATGTTTAAACAAAGAGAGGCTATCCAGACCTTACAAAGATGAGAG.. 1601
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 AspGlnIleThrGlyLysPheHisSerArgThrLeuPheGlnLeuLysG1 365
1602 .....ATGTCATTCCTCGAAGATCAAGATCAAGCCATCTAGCA 1635
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 nArgArgAsnValProMetLysProProAspGluAlaLysValGlyAspT 382
1636 CACTT...GACACAGAAAGAAATGAGACAGATTAACAGATGAGGCA 1682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 hrLeuGlyAspSerSerSerSerValLeuGluPheMetSerMetLys... 397
1683 CCAGGCGCTTGGACTCCAGTACTCAGACAGAGCTCATAGCGCTGAAGC 1732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
398 .....SerTyrProAspValSerValAspIleSerMet 408
1733 AGTAAATTTTGGCTTCCAGTTCGATGAGCTGCTGCTGAGAAAGCTG 1782
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 lLeuSerSerLeuGlyLysValLysLysGluLeuAspProAspAspSerH 425
1783 AAATATTCAGAAAGACTTGATGAGACACACAGATTCCTCAGGAGACT 1832
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 lSerLys.....AsnLeuAspGluThrThrLysLeuLeuLinsPheL 438
1833 CAGGAAGCCAGATGAACGTTTGAACACAGACCCCTGGGAGACATGAT 1882
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
439 HisGluAlaGlnAlaGlnArgGlyLysSerArgProSerSerAsnLeuSe 455
1883 CTGCTCTTGGTGCCCTCA 1901
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 rSerLeuSerAsnAlaSer 461

```

seq_name: /SIDS1/gcgdata/hold-geneseq/genesep-emb1/AA2000.DAT: AAB41780

seq_documentation_block:

ID AAB41780 standard; protein; 280 AA.

XX AAB41780;

AC AAB41780;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF1544 polypeptide sequence SEQ ID NO: 3088.

DE Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;

KW antiviral; antiparasitic; antiparkinsonian; immunotropic; neuroprotective;

KW anticonvulsant; osteoporotic; antidiarrheal; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;
thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

N-PSDB; AAC75989.

Novel nucleic acids and peptides derived from open reading frame X,
neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 2306-2307; 5507Pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytostatic; hepatotropic; vulnery;
antiproliferative; antiparkinsonian; nootropic; neuroprotective;
osteoplastic; anticoagulant; antithrombotic; immunosuppressive;
immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antiinflammatory; antibacterial; antiviral; antineoplastic;
antihypertensive; antianaemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an ORFX-associated disorder. The
nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antiinflammatory disease; to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 280 AA;

alignment_scores:

Quality: 476.00 Length: 311
Ratio: 2.441 Gaps: 5
Percent Similarity: 62.701 Percent Identity: 32.797

alignment_block:

US-09-687-230-1 x AAB41780 ..

Align seg 1/1 to: AAB41780 from: 1 to: 280

369 GAGAGCAGATTCCAGGGGAGAGGAGAGGAGAGGAGAGGAGACTTAA 418

```

|||||..... ||||| ..|||...|||
2  GILYSHISLEUSPASPGLUGLU..... ArgArgysArgylusgl 15
419  GGAGGATMAAAGAGGAGATGAGACCGGGGAGAAATGAGGACAGA. 467
    ||||| |||||..... |||||..... |||||..... |||||
15  UGLULYSARGLYSARGLYSARGLYSARGLYSARGLYSARGLYSARGLYS 32
468  ..AAGATCTCCAGTGCACGCCCTGTGAGTATGATGATGATGATGATG 515
    |||||..... |||||..... |||||..... |||||
32  IASPASPHEASPPROGLYLYSGLYSVALGLULVALGLUPROPPROASP 48
516  AAGCTCTCTCAAGACTCTTTAGCCAAAGAAAGAAAGTACAGACAGACC 565
    |||||..... |||||..... |||||..... |||||
49  ArgProValArgAlacysArgThrGlnGlnProGluMetGluArgThrH 65
566  CCTTCAAGAGCTTGTGATGATGAGACCAATGACGAGAAAGAAAGTAC 615
    |||||..... |||||..... |||||..... |||||
65  stIEGlnGlnLeuLeuGlnHisPheLeuArgGlnLeuGlnArgLysasp 82
616  CAAGTGTCTTCTTTTCATTTCTGTGACTGATTTTATGCTCTGCTGCTAC 665
    |||||..... |||||..... |||||..... |||||
82  roHisglYPhePheAlaPheProValThrAspAlaIleAlaProGlyTyr 98
666  TCCATGATCATTTAAACACCCCAATGATTTTACTACATGAGAAAGAGAT 715
    |||||..... |||||..... |||||..... |||||
99  SerMetIleIleLysHisPheMetAspPheGlyThrMetLysAspLysI 115
716  CAAGAACCAATGACTATGATGATGATGATGATGATGATGATGATGATG 765
    |||||..... |||||..... |||||..... |||||
115  eValAlaAsnGlnTyrLysSerValThrGlnPheLysAlaAspPheLys 132
766  TAAATGTACTAATGATGATGATGATGATGATGATGATGATGATGATG 815
    |||||..... |||||..... |||||..... |||||
132  euketCysAspAsnAlaMetThrTyrAsnArgProAspThrValTyrTyr 148
816  AAGCTGCAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 865
    |||||..... |||||..... |||||..... |||||
149  LysLeuAlaLysLysIleLeuHisAlaGlyPheLysMetLysSerLysgl 165
866  AAGAAAT..... CAGAGCTGAGAGAGAGAGAGTACGTACGTACGTAC 897
    |||||..... |||||..... |||||..... |||||
165  nAlaIleLeuLeuGlyAsnGlnAspThrAlaValGlnGlnProValProG 182
898  TCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 947
    |||||..... |||||..... |||||..... |||||
182  lValValProValGln..... CAGAGAGAGAGAGAGAGAGAGAGAGAG 187
948  TCACAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997
    |||||..... |||||..... |||||..... |||||
187  ..... 187
998  TGGAGATGCCGAAGCACAGCGCTTCAAGAGTCCACGAAGAATAATAAA 1047
    |||||..... |||||..... |||||..... |||||
188  ..... ValGlnThrAlaLysLysSerLysL 196
1048  AGAAGACAAAGATATGCTTGAAGATTAAGTTAA..... 1082
    |||||..... |||||..... |||||..... |||||
196  yProSerArgGlnValIleSerCysMetPheGlnProGlnGlnAla 212
1083  ..... AGCAATAATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
    |||||..... |||||..... |||||..... |||||
213  CysSerLeuThrAspSerThrAlaGlnGlnHisValLeuAlaLeuValS 229
1118  CCGCATCGTGAAGAGATCTGAGAGAAAGTGAACGAGCGCTTGTGACAA 1167
    |||||..... |||||..... |||||..... |||||
229  uHisAlaAlaAspGlnAlaArgAspArgLysAsnArgPheLeuProGly 246
1168  GTCACTGCGAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
    |||||..... |||||..... |||||..... |||||
246  lYsMetGlyTyrLeuLysArgAsnGlyAspGlySerLeuLeuTyrSer 262
1218  CTTTCACATCTGTGATCCCATTTGAGAGAG 1250
    |||||..... |||||..... |||||..... |||||

```

263 ValValaSnThrAlaGluProAsnAlaAspGlu 273

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA95881

seq_documentation_block:

ID AA95881 standard; Protein: 233 AA.

AC AA95881;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:18979.

KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX

PI Ota T, Isegaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI: 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

PS Claim 8; SEQ ID 18979; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

SO Sequence 233 AA;

alignment_scores:
Quality: 391.00 Length: 106
Ratio: 4.116 Gaps: 0
Percent Similarity: 89.623 Percent Identity: 65.094

alignment_block:

US-09-687-230-1 x AA95881 ..

Align seg 1/1 to: AA95881 from: 1 to: 233

549 GAAGTAGAACAGACACCCCTTCAAGAAAGCTTGAATCACTGATGAGCA 598

19 GlnAsnGluSerThrProIleGlnIleuLeuAspHisPheLeuArgI 35

599 ATTGCAGAGAAAGATCCAGTGGCTTCTTTGCAATTCCTGTGACGTAT 648

35 nleuGlnArgLysAspProHisGlyPhePheAlaPheProValThrAsp 52

649 TTATTCCTCCTGGCTACTCCATGATCATTTAAACACCCCATGGATTTAGT 698

52 IalIeAlaProGlyTyrSerMetIleIleLysHisPrometAspPheGly 68

699 ACCATGAAAGAAAGATCAAGAACATGACTTTCAGTCCATAGAGACT 748

69 ThrMetLysAspLysIleValAlaAsnGluTyrLysSerValThrGluPh 85

749 AAAGATCACTTCAACCTAATGTGTACTAATGCCATGATTTACATTAAC 798

85 elysAlaAspPheLysLeuMetCysAspAsnAlaMetThrTyrAsnArgp 102

799 CAGAGACCATTTATTTATTAAGCTGCAGAAAGAGCTGTTGACATCAGAGATG 848

102 roAspThrValTyrTyrLysLeuAlaLysLysIleLeuHisAlaGlyPhe 118

849 AAAATTCCTAGCCAGAA 866

119 LysMetMetSerLysGln 124

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAU16626

seq_documentation_block:

ID AAU16626 standard; Protein: 718 AA.

XX

AC AAU16626;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 1579.

KM Human; immunosuppressive; antiarthritic; antirheumatic;

KM cytototoxic; cardiant; vasotropic; cerebroprotective; nootropic;

KM neuroprotective; antibacterial; virocidic; fungicide; opthalmological;

KM vulnerray; secreted protein; rheumatoid arthritis;

KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

KM cerebrovascular disorder; cerebral ischemia; angiogenesis;

KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;

KM corneal infection; wound healing; epithelial cell proliferation;

KM skin aging; food additive; preservative; antiproliferative.

XX

OS Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01341.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232387.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237049.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-488783/53.

N-PSDB: AAS26613.

New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives -

Claim 11; SEQ ID NO 1579; 980bp; English.

The invention relates to isolated nucleic acid molecules and their
encoded secreted proteins. The nucleic acids and proteins are used to
prevent, treat or ameliorate a medical condition in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. They
are also used in diagnosing a pathological condition or susceptibility
to a pathological condition. Antibodies to the proteins can also


```

601 TGCAGAGAAAAGATCCAGTGTCTTTTCATTTCTGCTGACTGATTTT 650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 euglnasplysaspproalaarglilephealaglprovalserleuyls 146
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
651 ATTGCCCTGGCTACTCCATGATCATTAACNCCCAATGGATTTTGTAC 700
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
147 gluvalprasptryleuasphtllyshlsprometalsphhealath 163
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
701 CATGAAGAAGAAAGATCAGAACATGACTATCATGCTCATAGAACTAA 750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 rmetarglysargleugluaglnclytrylryasnlhehlsclupheg 180
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
751 AGCATTAAGTCAAGTAATGTGTACTAATGCCATGATTTACAAATAACA 800
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
180 lnglunaspheaspleulleileaspasnlysmetylstryasnalaarg 196
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
801 GAGACCATTTATATTAAGCTGCAAGAAGCTTTGCTGCTGACTGAGATGA 850
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
197 AAPTThValPheTyrArgAlaAlaValArgLeuAlaArgAspGlnIylva 213
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
851 AATTTTACCCAGGAAGAATTCAGAGCTGAAGCAGACATAGACTTCA 900
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
213 lvalleuarglnalaarg..... 219
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
901 TGGCTGACTTGCAGAAACTCGAAGACAGAAAGATGAAACAGACACTCA 950
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
219 ..... 219
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
951 CAGAGTGGGAGAGAGAGGCTGCTGCGAGAGAGAGAGAGAGCTCTGG 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
220 ..... 226
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1001 AGATGCCAGAGCA.....CACGCTTCAAGAGCTCCAGCAAGAA 1041
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 yleuglunlAlaSerGlyMetHisleuProgluArgproAlaAlaAlap 243
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1042 ATAAAAAG.....AAGACAAAGATRTGCTTGAAGATTAAGTTT 1079
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
243 roArgTArgProPheSerTrpGluAspValAspArgleuLeuaspProAla 259
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1080 AAAAGCAATTAATTTA.....GAGAGAGAGAGAGAGAGAGAGAGCG 1120
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
260 AsnAlArgAlaHisleuGlyleuGlnleuAlaArgGlnleuLeuaspHe 276
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1121 CATC.....GTGAAGCAATCTGAGAGAAAGCTGACAGC. 1154
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
276 tleuaspLeuThrCysAlaMetIysSerSerGlySerArgserIylvaArg 293
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1155 ..CGGCTTGTGAACAGTCAAGTGCATTTGAAGAAGAAACCAGATGGA 1202
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
293 lAlysleuLeuLysGlnIleAlaLeuAlaArgAsnLys..... 306
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1203 ACAAGAGAGCTGGGACTTCCATCTGTGATGCCATGTTGGAGAGGCC 1252
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
307 .....LeuSerGlnGlnHisSer..GlnProLeuProThylcylpr 319
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1253 AGGCTACTGCTGCTGAGACTGGGAATGACAACTGGAAGACTTCACTC 1302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 ogly..... 320
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1303 GAGTGAATTAATTTCAGAGGGGTTCAAGAGAGAT 1334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 .....LeuGlnGlyPheGlnGluAsp 327
    ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
seq_name: /SIBS1/gcdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:AA39231
seq_documentation_block:
ID   AAM39231 standard; Protein; 1058 AA.
XX
AC   AAM39231;
XX
DF   22-OCT-2001 (first entry)

```

```

XX DE Human polypeptide SEQ ID NO 2376.
XX DE
XX XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.
XX OS Homo sapiens.
XX FN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Weinman R, Xu G, Xue A, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AAI58387.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 4; SEQ ID NO 2376; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cyostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SO Sequence 1058 AA;

```

alignment_scores:
 Quality: 262.50 Length: 485
 Ratio: 1.156 Gaps: 18
 Percent Similarity: 46.804 Percent Identity: 23.918

alignment_block:
 US-09-687-230-1 x AAM39231 ..
 Align seg 1/1 to: AAM39231 from: 1 to: 1058

```

17 CGGCCGGCGCGCGCGCGCGCTGCTCG..... 46
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

390 ProProGlycysThrArgArgProLeuAsnIleTyrGlyAspValGluLeu 406
47 .CGCGCGGGGGTCTCGCGGGCCCCCGCTCCGCGCTCGCTCGCTGGCCCC 95
   :::::|||||
406 tlysAsnGlyValCysArgLys.....GluSerSerValIlysthrValA 421
   :::::|||||
96 GGACCGGAACCGCGCGCGCTGGCGCTGGCGGGGGCGGGCGAC 145
   :::::|||||
421 rgsSerThrSerLysValArgLysValLysAlaLysLysAlaLeu 437
   :::::|||||
146 CGGGGGCGCGCTCGGACATGGGCAAGAACACAGAACACAGTCGGACA 195
   |||
438 AlagLupProCys..... 441
196 AACACCTCTACGAGAGCTATGT..... 217
   :::::|||||
442 .AlaValLeuProThrValCysAlaProTyrIleProProGlnArgLeuA 458
   :::::|||||
218 .AGAGAGCCCTTGAAGCTGGTCTCAAGTAGAGG.....GAA 256
   |||
458 snArgIleAlaAsnGlnValAlaIleGlnArgLysLysGlnPheValGln 474
   :::::|||||
257 CGAAGTCACCGA...ACTCTCACGGGCGCTCGGGCGGCGACTCCAGCC 303
   |||
475 ArgAlaIleSerTyrTrpLeuLeuLysArgLeuSerArgAsnGlyAlaPr 491
   :::::|||||
304 TCCTCGAGACCAAAACGATCATGACAAACACAGACAGAAACGGGAAA 353
   |||
491 OleuLeuArgArgLeuGlnSer..SerLeuGlnSerGlnArgSerSerGln 507
   :::::|||||
354 AAGAGAAAGAAAGCAGAGAACGACATTCCAGGGGAAACAAAAGGGGAAA 403
   :::::|||||
508 GlnArgGlnAsnArgLysGlnUmeLysAlaLysLysLysLeuLysTy 524
   :::::|||||
404 ACGGAGAAAGTTAAGGAGGATAAAGAG..... 434
   :::::|||||
524 rTrpGlnArgLeuArgHisAspLeuGlnArgAlaArgLeuLeuIleGlu 541
   :::::|||||
435 .CGAGTCGAGACCGGGTGAGATGAGCGCAAGAAAGATCTCCAG 479
   |||
541 euleuArgLysArgGlnLysLeuLysArgGlnGlnValGln 557
   :::::|||||
480 TCACACCGCCCTGAGATTAGACTTCCTCAGAACGCTCAGACAG 529
   :::::|||||
558 ValAlaMetGluLeuArgLeu..... 564
530 CTCCTTAGCCAAACAGAGAGTAGAACAGACCCCTTCAGAGAGCTT 579
   |||
565 .....ThrProLeuThrValLeuL 571
580 TGAATCACTGATGAGACAATTCGAGAGAAAGATCCAGTGTCTTCTT 629
   |||
571 eulArgSerValIleuAspGlnLeuGlnAspLysAspProAlaArgIlePhe 587
   :::::|||||
630 TCATTTCTCTGACTGATTTTATGCTCCCTGCGCTCCATGATCATTA 679
   :::::|||||
588 AlaGlnProValSerLeuLysGlnValProAspTyrLeuAspHisIleLys 604
   :::::|||||
680 ACACCAATGGATTTTAGTACATGAAAGAAAGATCAAGAACATGAGCT 729
   |||
604 snHisProMetSerPheAlaThrMetArgLysArgLeuGlnAlaGlnGlyT 621
   :::::|||||
730 ATCAGTCCATGAGAACTAAAGGATTAATCAACTAATGTGTAATAAT 779
   |||
621 yLysAsnLeuHisGlnPheGlnGlnAspPheAspLeuIleIleAspAsn 637
   :::::|||||
780 GGCATGATTTTCAATAAACGAGACCATTTATTAAGCTGCAAGAA 829
   :::::|||||
638 CysMetLysTyrAsnAlaArgAspThrValPheTyrArgAlaAlaValArg 654
   :::::|||||
830 GCTGTTCACATCAGAGAAATGAAATTTCTTACGCAAGAAATTCAGAGCC 879
   |||
654 gLeuArgAspGlnGlnGlyValValLeuArgGlnAlaArg..... 667

```

```

880 TGAAGCAGACATAGACTTTCATGCTGACTTGCAGAAAACCTCGAAACAG 929
667 ..... 667
930 AAGATGGAACAGACACCTTCACAGAGTGGGAGACGAGGCTGTGTGCA 979
667 ..... 667
980 GAGAGAGAGAGAGACTCTGAGATGCCGAGACA.....CAGCGCT 1020
   |||
668 .ArgGlnValAspSerIleGlyLeuGlnGlnAlaSerClyMetHisLeuP 684
1021 TCAAGACTCCCGCAAGAAATTAAG.....AAGACAA 1058
   :::::|||||
684 rGlnArgProAlaAlaAlaProArgArgProPheSerTrpGlnAspVal 700
   :::::|||||
1059 GATATGCTTGAAGTATAGTTTAAAGCAATTAATTTA.....GAGAG 1099
   |||
701 AspArgLeuLeuAspProAlaAsnArgAlaHisIleuGlyLeuGlnGln 717
   :::::|||||
1100 AGAGCAGAGACAGCTTGACCGCATC.....GTGAAAGGAT 1134
   :::::|||||
717 nLeuArgGlnLeuLeuAspMetLeuAspLeuThrCysAlaMetLysSer 734
   |||
1135 CTGGAGGAAAGCTGACCCAGG...CGGCTTGTGAACACTCAGTGGCAATTT 1181
   |||
734 ergLysSerArgSerLysArgAlaLysLeuLeuLysGlnIleAlaLeu 750
   :::::|||||
1182 GAAGAGAAAGAACCATGATGGAACAACAGACGTGGGACTTCTCCATCTGT 1231
   |||
751 LeuArgAsnLys.....LeuSerGlnGlnHisSer.. 760
1232 GGATCCCATTTGTAGGAGAGCCAGGCTACTGCTGTGTGAGACTGGAAATGA 1281
   :::::|||||
761 .GlnProLeuProThrGlyProGly..... 768
1282 CAACTGGAAGACTTCAGCTGTGAGCTGAATCTTTGCAAGGGGTTCAAGAG 1331
   |||
769 .....LeuGlnGlyPheGlnGln 774
1332 GAT 1334
   |||
775 Asp 775

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:49:37 : Search time 17.51 Seconds
(without alignments)
1302.445 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073
Sequence: 1 MGKKHKHSDKHLYEYVE.....PQNMICLLGPSEKCLLNK 589

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258.5	8.4	1058	BRD1_HUMAN	O95696 homo sapien
2	248	8.1	1214	BRE3_HUMAN	Q91404 homo sapien
3	235.5	7.7	1214	BRE1_HUMAN	P55201 homo sapien
4	210.5	6.8	2442	CBP_HUMAN	O92793 homo sapien
5	210	6.8	1586	SN22_HUMAN	P15317 homo sapien
6	208	6.8	2068	T2D1_DROME	P51123 drosophila
7	199.5	6.5	1332	SP77_YEAST	P35177 saccharomyc
8	198	6.4	2441	CBP_MOUSE	P45481 mus musculu
9	195	6.3	733	YC40_HUMAN	Q91110 homo sapien
10	190.5	6.2	2414	P300_HUMAN	O09472 homo sapien
11	190	6.2	1359	ATRX_CAREL	Q91760 caenorhabdi
12	187	6.1	1647	SN24_HUMAN	P15352 homo sapien
13	181.5	5.9	1872	T2D1_HUMAN	P21675 homo sapien
14	179.5	5.8	801	BRD2_HUMAN	P35440 homo sapien
15	167.5	5.5	686	BDF1_YEAST	P35817 saccharomyc
16	165	5.4	1982	CHDM_DROME	O97159 drosophila
17	164.5	5.4	1362	BRD4_HUMAN	O60885 homo sapien
18	163	5.3	1638	BRM_DROME	P25439 drosophila
19	162	5.3	1790	USO1_YEAST	P25386 saccharomyc
20	161.5	5.3	728	TRDN_HUMAN	Q13061 homo sapien
21	157	5.1	726	BRD3_HUMAN	P15059 homo sapien
22	156.5	5.1	678	GARP_PLAUF	P13816 plasmodium
23	156	5.1	1959	MYH9_CHICK	P14105 gallus gall
24	155	5.0	1920	PCNT_MOUSE	P48723 mus musculu
25	151	4.9	2116	MYH2_DICDI	P08799 dictyostell
26	150	4.9	1972	MYHB_HUMAN	P35749 homo sapien
27	149.5	4.9	705	TRDN_RABIT	Q28820 oryctolagus
28	149.5	4.9	1087	AKA9_RABIT	Q28638 oryctolagus
29	149	4.8	439	GCN5_YEAST	Q03333 saccharomyc
30	149	4.8	1433	REST_CHICK	O42184 gallus gall
31	148	4.8	1972	MYHB_RABIT	P35748 oryctolagus
32	147.5	4.8	1359	STH1_YEAST	P32591 saccharomyc
33	146.5	4.8	1976	MYHA_BOVIN	Q27991 bos taurus

34	146.5	4.8	2464	1	MAPB_MOUSE	P14873 mus musculu
35	146	4.8	2805	1	MAPA_HUMAN	P78559 homo sapien
36	145.5	4.7	1972	1	MYHB_MOUSE	O08638 mus musculu
37	145.5	4.7	2653	1	CENE_HUMAN	O02224 homo sapien
38	145	4.7	1938	1	MYH6_MOUSE	O02566 mus musculu
39	145	4.7	2295	1	MDR9_HUMAN	O9ns16 homo sapien
40	144	4.7	829	1	TOP1_XENLA	P41512 xenopus lae
41	144	4.7	1938	1	MYH6_RAT	P02563 rattus norv
42	143.5	4.7	717	1	HS9B_BRUPA	O61998 brugia paha
43	143.5	4.7	767	1	TOP1_CRIGR	O07050 cricetus
44	143.5	4.7	1961	1	MYH9_RAT	O62812 rattus norv
45	143	4.7	1225	1	SMC1_YEAST	P32908 saccharomyc

ALIGNMENTS

```

RESULT 1
ID BRD1_HUMAN STANDARD: PRT; 1058 AA.
AC O95696;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bromodomain-containing protein 1 (BR140-like protein).
GN BRD1 OR BRL OR BRPF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20071128; PubMed=10602503;
RA McCullagh P., Chaplin T., Meerabux J., Grenzelias D., Lillington D.,
RA Poulson R., Gregorini A., Saha V., Young B.D.;
RT "The cloning, mapping and expression of a novel gene, BRL, related to
RT the A10 leukaemia gene."
RL Oncogene 18:7442-7452(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Hunt A.;
RL Submitted (MUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 1 PWWP DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: AF005067; AAF34320.1; -.
EMBL: Z98885; CAB11574.1; -.
EMBL: 604589; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR000313; PWWP.
DR Pfam: PF00439; Bromodomain; 1.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF00855; PWWP; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00293; PWWP; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50812; PWWP; 1.
KW Nuclear protein; Zinc-finger; Bromodomain.

```

FT ZN_FING 214 264 PHD-TYPE.
 FT DOMAIN 579 649 BROMODOMAIN.
 FT 929 1012 PMP.
 SO SEQUENCE 1058 AA; 119519 MW; 6E7B07E8A030E104 CRC64;

Query Match 8.4%; Score 258.5; DB 1; Length 1056;
 Best Local Similarity 23.5%; Pred. No. 2e-07;
 Matches 109; Conservative 62; Mismatches 136; Indels 157; Gaps 15;

OY 11 DKHLYEYKPKLVKXGNVTELTSGSSGHSLSFEDKNDHKHDKRRKKKKKE 70
 DB 386 DVHTPGCTRRRLNLYGDV-----EMKNGVCRRKSSVATVS-----TSKYKKKKKKAK 434
 OY 71 K-----QIGEEKRRRRKRRKEDKK----- 91
 DB 435 KALAECAVLPTVCAPYIPQRLNRIANQVAILORKQFVBRASHYLLKRLSNGAPLIR 494
 OY 92 -----RDRRENEAEKDLQCHAVRLDLPEKPLTSSLAQE-----E 130
 DB 495 RLQSSLSQSRSSQGRNDEEMKAKKELKYWQRLRDLERARLLIELKREKLREQYK 554
 OY 131 VEG-----TLQELNLMQLQKRDSPAFSPVPTDFIAGYSMIITKHPPDESTMKE 183
 DB 555 VEGVAMELRLLPTVLLRSVLDLQDKDPARIFAPQVSLKEVPDYLDHDKHPDFTMKR 614
 OY 184 KIKNDYOSIEELKDNFKLMCTNAMIYNNKPEIYYKANKLLHSGMKILISOERISQSLKOS 243
 DB 615 RLEAGQYKRLHEFEEDFDLLIDNCMKYNADYFYRAAVRLRQGGVVLROAR----- 667
 OY 244 IDPMADLQTRKQKGTDTISQSGEDGCGWOREDSGADA---NAFKSPSKENK--- 296
 DB 668 -----REVDSIGLEASGSMHLPBPAPAPRRPFSSM 697
 OY 297 KDKDMLDEKFSNNL---FREDFOLDRI-----VKESGKLTFR-RLVNSOCEREPKRPDG 347
 DB 698 EDVDRLLDANRNLHLEQRLKRLMDLITCMKSSGSRKAKLLKKEIALIRNK--- 754
 OY 348 TTTGLLHPVDPIVGEPCYCLVRLGTTGRLOSQVNTLOGFKED 391
 DB 755 ---LSQOHS-QPLPTGPG-----LGSFEED 775

RESULT 2
 BRP3_HUMAN STANDARD: PRT: 1214 AA.
 AC 09UD4;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Bromodomain and PHD finger-containing protein 3 (Fragment).
 GN BRP3 OR KIAI1286.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 RA Ohara O.;
 RT Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.;
 RL DNA Res. 6:337-345(1999).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 PMP DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB031112; BAAB6600.1; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR001965; PHD.
 DR InterPro; IPR000313; PMP.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF00628; PHD; 1.
 DR Pfam; PF00855; PMP; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00293; PMP; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50812; PMP; 1.
 KW Zinc-finger; Bromodomain.
 FT ZN_FING 223 271 PHD-TYPE.
 FT DOMAIN 415 441 GLU-RICH.
 FT DOMAIN 615 685 BROMODOMAIN.
 FT 1085 1168 PMP.
 SO SEQUENCE 1214 AA; 136598 MW; CA490810622109CD CRC64;

Query Match 8.1%; Score 248; DB 1; Length 1214;
 Best Local Similarity 20.8%; Pred. No. 9.2e-07;
 Matches 129; Conservative 95; Mismatches 220; Indels 176; Gaps 20;

OY 22 PLKVLKVGNGNVTELTSGSSGHSLSFEDKNDHKHDKRRKKKKKKKQIPGEGKGR 81
 DB 481 PLVAVQIISYSLNKICSLT-----FQKNGQFQRLHNYLLKRLQANGVPLRLRLHS 534
 OY 82 RRRVKEKKRRDRKVENAEKDLQCHAVRLDLPEKPLTSSLAQEYV--EQPLQEA 139
 DB 535 HLQSRNABQRODEKTSVAKKELKYWQRLRDLERARLLIELKREKLREQYKVOOA 594
 OY 140 -----LNQLMRQ-----LQKRDSPAFSPVPTDFIAGYSMIITKHPPDFTMKETKN 187
 DB 595 AMELIEMPPNVLLRTLLDLQEKDPAHFAEPVNISEVPDYLEFISKPDSTMRKLES 654
 OY 188 NDYOSIEELKDNFKLMCTNAMIYNNKPEIYYKANKLLHSGMKILISOERISQSIDFM 247
 DB 655 HLYRTLFEFEEDFNLIYTCMKYNAKDTIFHRAAVRLDGLGAILRHARQENIGYD-- 712
 OY 248 ADLQTRKQKGTDTISQSGEDGCGWOREDSGD---AEAHAFKSPSKENKAKDKMLD 304
 DB 713 -----PERGTHLPESPKLDFRFSWEDVNTLLIPENRAHLSPEVO----- 753
 OY 305 KFKSNLLEKDFOLDRI---VKESGKLTFR-RLVNSOCEREPKRPDGTTLGLLHPVDPIV 361
 DB 754 -----LKEHLKLDLVSMRSGARTRVRLIRREINLRK-----LAQPPP-- 797
 OY 362 GEPCYCLVRLGTTGRLOSQVNTLOGFKEDKRNKYTPVLYLYNGYSVAPHYDSTFANI 421
 DB 798 -----QPSLNKTVSNGLPAGQGD-AAVLQALQEE 829
 OY 422 SKDDSLITYSGEDSLPDSFIHFPLATCODYRYVADSLIDVLTGKGHSRTLOEMEM 481
 DB 830 PEDDGG-----RDSKLTTP-----PTLEPTGPAISLSEGE- 860
 OY 482 SLPEDEGHTRLDTGKEMQITEVEPPGR-----LDS 513
 DB 861 SPPPP-----PLTKPINDSKRPPKFLKRAVEDELLKSPLOLGNHPTLRLSD 910
 OY 514 STQDRLLAKAVTNFQVPEVDFSEAEITQK-----KIDET-TLLBELQF----- 559
 DB 911 NQINRLSLAAPPTAGTPLSGVGRRTSVLFFKRAKKGVKLQRPDVLNLEGGEDHGAAGSPA 970
 OY 560 -----AQNERLSTRPPQNMIC 575

```

Db      971 SPATSEERHSKRPRSRSC 990

RESULT 3
BREFL_HUMAN STANDARD; PRT; 1214 AA.
AC      PS5201; O9UHI0;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DE      15-OCT-2001 (Rel. 40, Last annotation update)
DE      Pteregrin (Bromodomains and PHD finger-containing protein 1) (BR140
DE      protein).
GN      BRP1 OR BR140.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94161726; PubMed=7906940.
RA      Thompson K.A., Wang B., Argaves W.S., Giannotti F.G., Schranck D.P.,
RA      Rosslahdt E.;
RT      "BR140, a novel zinc-finger protein with homology to the TAF250
RT      subunit of TFIID."
RL      Biochem. Biophys. Res. Commun. 198;1143-1152(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Hu S.N., Dong W., Zeng Y.X., Yu J., Yang H.M.;
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: UNKNOWN. POSSIBLE TRANSCRIPTION ACTIVATOR.
CC      -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC      -1- TISSUE SPECIFICITY: HIGH LEVELS IN TESTIS.
CC      -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC      -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC      -1- SIMILARITY: CONTAINS 1 PMWP DOMAIN.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M91585; AAB02119.1; -.
DR      EMBL; AF176815; AAF19605.1; -.
DR      MIM; 602410; -.
DR      InterPro; IPR001487; Bromodomain.
DR      InterPro; IPR001965; PHD.
DR      InterPro; IPR000313; PMWP.
DR      InterPro; IPR000822; Znf-C2H2.
DR      Pfam; PF00439; bromodomain; 1.
DR      Pfam; PF00628; PHD; 1.
DR      Pfam; PF00855; PMWP; 1.
DR      PRINTS; PRO0503; BROMODOMAIN.
DR      SMART; SM00297; BROMO; 1.
DR      SMART; SM00249; PHD; 2.
DR      SMART; SM00355; Znf-C2H2; 1.
DR      SMART; SM00633; BROMODOMAIN_1; FALSE_NEG.
DR      PROSITE; PS50812; BROMODOMAIN_2; 1.
DR      PROSITE; PS50814; PMWP; 1.
DR      PROSITE; PS50028; ZINC_FINGER_C2H2_1; 1.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW      Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW      Zinc-finger; Bromodomain.
FT      ZN_FING 21 47 C2H2-TYPE.
FT      ZN_FING 273 323 PHD-TYPE.
FT      ZN_FING 386 400 C4-TYPE.
FT      DOMAIN 645 715 BROMODOMAIN.
FT      DOMAIN 1085 1168 PMWP.
FT      CONFLICT 299 299 E -> A (IN REF. 2).

```

```

FT      CONFLICT 729 729 V -> L (IN REF. 2).
SQ      SEQUENCE 1214 AA; 137542 MW; C530CD2F3083A53D CRC64;

Query Match
Best Local Similarity 26.3%; Pred. No. 4,7e-06;
Matches 82; Conservative 58; Mismatches 129; Indels 43; Gaps 10;

QY      82 RRRYEDKKRRDRVRNEAE-----KDIQCHAPRLDLPPEKPLTSSLANOE-----129
DB      558 RLQTHLQSORNCQVGRSEDKNWLKLEQLSKWRLHDLERARLLVETLRREKREKRE 617
QY      130 -----EVEQPTLOEALNQLMROQLRKDPSPAFSEFPYPTDFIAPYSIIKHHPMFST 180
DB      618 TIKYQIAMEQLTPPLILRLKLTLEQLQEKDTGNIFSEPPVLPSEVPYLDHKKPMDQFT 677
QY      181 MKEKIKNNDOSIELKDNFKLMCTNMIYNNKPTIYKAKKLLHSGMKLTISOERQSL 240
DB      678 MKQNLVARYRLNPDFFEDNLIYSNCLKYNADTIYRAAVRLREGGAVVQARQAE 737
QY      241 KQSIDPADIQKTRKQKQDTSQSGEDGCGWQREDSGDAEAHAFKSPSKENKKRKDK 300
DB      738 KMGIDFETGMHHPISLA-GDEATHTEDAA--EERLVLLNQKHL--PYEQLK----787
QY      301 MLEDKFSNNIERE---QEQLDRIYKESGKLTIRLYNSQCEFERKPD--GTTTGLLH 355
DB      788 LLERLDEVNASKOSVGRSRRAKMKIKEMTALRRKLAH-QREGRDGPERRGPSRSSGLT 846
QY      356 PVDPVGEPCYC 367
DB      847 P-----HPAAC 852

RESULT 4
CBP_HUMAN STANDARD; PRT; 2442 AA.
ID      092793; Q16376; Q00147;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      01-MAR-2002 (Rel. 41, Last annotation update)
DE      CREB-binding protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97385172; PubMed=9238046.
RA      Sobulo O.M., Borrow J., Tomek R., Reshlmi S., Harden A.,
RA      Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA      Zelezniik-Ie N.J.;
RT      "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT      acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL      Proc. Natl. Acad. Sci. U.S.A. 94;8732-8737(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97321049; PubMed=9177780;
RA      Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA      Lushikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA      Doggett N.A., Peters D.J.M., Breuning M.H.;
RT      "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT      of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT      16p13.3.";
RL      Genomics 42;96-144(1997).
RN      [3]
RP      SEQUENCE OF 1-405 FROM N.A.
RX      MEDLINE=96376968; PubMed=8782817;
RA      Borrow J., Stanton V.P., Andersen J.M., Becher R., Behm F.G.,
RA      Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA      Housman D., Mittleman F., Volinia S., Watson A.E., Housman D.E.;
RT      "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT      a putative acetyltransferase to the CREB-binding protein.";

```

RL Nat. Genet. 14:33-41(1996).
 CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AGUMENTS
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CC CAMP-RESPONSIVE GENES.
 CC -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;p13.3)
 CC INVOLVING CBP AND MLL.
 CC -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
 CC SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
 CC ABNORMALITIES, BROAD THUMBES, BROAD BIG TOES, MENTAL RETARDATION
 CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U47741; AAC51770.1; -
 DR EMBL; U85962; AAC51331.1; -
 DR EMBL; U89354; AAC51339.1; -
 DR EMBL; U89355; AAC51340.1; -
 DR MIM; 600140; -
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003101; KIX.
 DR InterPro; IPR00197; ZNF_finger.
 DR InterPro; IPR000433; ZNF_ZZ.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02172; KIX; 1.
 DR Pfam; PF02135; ZF-TAZ; 2.
 DR Pfam; PF00569; Z2; 1.
 DR PRINTS; PRO0303; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00291; ZNF_ZZ; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS00014; BROMODOMAIN_2; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS01355; ZF_ZZ_2; 1.
 DR Transcription regulation; Nuclear protein; Activator; Bromodomain;
 KW Chromosomal translocation; Zinc-finger.
 FT ZN_FING 1701 1744
 FT DOMAIN 363 430 CYS/HIS-RICH.
 FT DOMAIN 452 683 CREB-BINDING.
 FT DOMAIN 1103 1175 BROMODOMAIN.
 FT DOMAIN 1061 1064 POLY-GLU.
 FT DOMAIN 1199 1487 CYS/HIS-RICH.
 FT DOMAIN 1555 1562 POLY-GLU.
 FT DOMAIN 1675 1849 CYS/HIS-RICH.
 FT DOMAIN 1943 1948 POLY-PRO.
 FT DOMAIN 1967 1970 POLY-GLN.
 FT DOMAIN 2081 2085 POLY-GLN.
 FT DOMAIN 2199 2216 POLY-GLN.
 FT DOMAIN 2245 2248 POLY-GLN.
 FT DOMAIN 2297 2300 POLY-GLN.
 FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
 FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
 FT CONFLICT 1770 1770 V -> L (IN REF. 2).
 FT CONFLICT 1789 1789 N -> F (IN REF. 2).
 FT CONFLICT 1812 1812 T -> P (IN REF. 2).
 SO SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

QY 36 ELSTGSSGHDSLSLEFDKND---HDKHKDRKKRKKKGGKQIPGEEKRRRRVKED---- 88
 Db 982 EFNSSQPGPDVPLKEMKTQAEDEPD-----PGESKGEPRSEMEEDLQG 1028
 QY 89 ---KKRDRRVRNEAKKDLQCHAPVRLDLPPEKPLTSLAKQEBVNGT----- 134
 Db 1029 ASQVKEETIAEKKSE-----PKEVD--EKKPEKVAYKEEBSSSNGTSSQSPSQ 1079
 QY 135 -----PLOGALNQLMRQLRKDPASAF--FSPVTDPI--APGSMITIKRPMDFSTWK 182
 Db 1080 PRKIKPPELRQALPPTLEALYRQPESLPPRPQVDPQLGIPDFYIVKMPMDLSTIK 1139
 QY 183 EKIKNDYOSIEBKNFELKTNMITYKPEITYYKAKKLLHSKMLTSQRIOSLQK 242
 Db 1140 RKLIDGQYDEPQYDDVDMFNMAWLYNRKTSRYKFKPSKL---AEVFEDE-IDPVWQ 1194
 QY 243 SLDFEMA-----DLQTKRKQGDGNTDSGSGDGC-----WQEREDSGA 282
 Db 1195 SLGYCCGRKYERSPQTLCTYQKQLCTIPRDAAYTSQNRHFCEKCFTEIQGENVTLLGD- 1253
 QY 283 EAHAFKSPSEKNNKKDKDMLIEDKFSNNLE-----REDEQL----- 318
 Db 1254 -----DPSQPGTTLSKQFEKK-KNDTLDPPFPVDCRCKGRKMQICVLHYDIIMPSEGF 1306
 QY 319 ---DRIVKESG-----GKTLRLVNSQCFEERRKPDGTTLLGLHPVDPIV 361
 Db 1307 VCDNCLKTKGRPRKKNKFSAKRLQTTRLGNHLEDVKNKELRRQN-----HP----- 1352
 QY 362 GEPGYCLVRLGWTGRLGSGVNTLGGFKEDKRNKTPVY---YLWGPYSYAPH-YDST 417
 Db 1353 -EAGEYFVRVAVS-----DKTEVAFGKSKSFVDSGKSESFPYRTKAL 1396
 QY 418 FANISKDSDLIY-----STYGEDSDLPD-----FSIHEFLATC---ODPYVWAD 461
 Db 1397 FAFEEIDGVDFCFGMHVGEGSDCPPWTRRVYISLYDSIHFFPRCLRTAVYHILLG 1456
 QY 462 SLIDVLTGK---GHSKTLQEMKSLPDEGG---HTRLDITGKMEQITVEYPPGRD- 512
 Db 1457 YLEVYAKRLGYVTGHT-----WACPPSGDDYIFCHCPPD-----QKIPKRRLOE 1501
 QY 513 -----SSTQDRLIALKAVTNGVVEVDSSEAEIFOKKIDE 549
 Db 1502 WYKKMLDKFAERIILHDYDIKQATREDLISAKELPYF-----EGDFWNVLEE 1551
 QY 550 TTRLRLLEQAOQNER 564
 Db 1552 S---IKELQEEBER 1563
 RESULT 5
 SN22_HUMAN STANDARD; PRT; 1586 AA.
 AC P51531;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Possible global transcription activator SNF2L2 (SNF2-alpha).
 GN SMARCA2 OR SNF2L2 OR BRM OR SNF2A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94038910; PubMed=8223438;
 RA Muchardt C., Yaniv M.;
 RT "A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and
 RT Drosophila brm genes potentiates transcriptional activation by the
 RT glucocorticoid receptor.";
 RL EMBO J. 12:4279-4290(1993).
 RN [2]
 RP SEQUENCE FROM N.A.

Query Match 6.8%; Score 210.5; DB 1; Length 2442;
 Best Local Similarity 21.2%; Pred. No. 0.0003;
 Matches 143; Conservative 80; Mismatches 213; Indels 239; Gaps 32;


```

OY 156 FSPVLTGFIAGYQSMILKHPDMFTSMKEKIKINDNOSIEELDKONFKLMCTNMIYKRPET 215
Db 1424 IQLESKREL-PEYELLIRKPDPDFKKIKERIRNNHRYRSLGDEKRYMLLCHNAQFNLNLS 1482
OY 216 IYKAAKKLLHSGMKILISOERIOSKSIDFMADLQTRKROKDTGTONSOGEDGCMORE 275
Db 1483 QIYEDSIYL-----OSV-FKSARQIKAEFEESDESDSNEBE-----EEE 1519
OY 276 REDSGDAEAAHAFKSPSKENKKKKDKMLMEDRKSS-----NNLREGEQJDRIVKE 324
Db 1520 DEESESESEAKSVKVKIKLNKKDKGRDKGKKRRPNNGKAKPVAVSDPDSDEQDERQSE 1579
OY 325 SGG 327
Db 1580 GSG 1582

RESULT 6
T2D1_DROME T2D1_DROME STANDARD; PRT; 2068 AA.
AC P51123:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIIID 230 kDa subunit (TAFTII-230)
DE (TAFTII250) (TBP-associated factor 230 kDa) (P230).
GN TAR250.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 63-75 AND 540-546.
RP MEDLINE=93279463; PubMed=8504928;
RX Kokubo T., Gong D.-W., Yamashita S., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Drosophila 230-kD TFIIID subunit, a functional homolog of the human
RT box-binding subunit of TFIIID."
RL Genes Dev. 7:1033-1046(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL FOR
CC PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES DNA-
CC BINDING ACTIVITY. IS A NEGATIVE REGULATOR OF THE TATA BOX-BINDING
CC ACTIVITY OF TBP.
CC -1- SUBUNIT: TF2D IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -1- SIMILARITY: TO HUMAN TAFII-250 (CCG1). SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAFI145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S61883; AAB26991.2; -.
DR TRANSFAC; T02119.
DR FlyBase; FBgn0010355; Taf250.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001878; ZnF_CCHC.
DR Pfam; PF002178; AT_hook; 1.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.

```

DR SMART: SM00384; AT_hook; 1.
 DR SMART: SM00297; BROMO; 2.
 DR SMART: SM00343; ZNF_C2HC; 1.
 DR PROSITE: PS00633; BROMODOMAIN 1; 2.
 DR PROSITE: PS50014; BROMODOMAIN 2; 2.
 KW Bromodomain: Nuclear protein; DNA-binding; Cell cycle; Repeat;
 KM Transcription regulation; Phosphorylation;
 FT DNA_BIND 1247 1360 HMG BOX (POTENTIAL).
 FT DOMAIN 1445 1451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1490 1560 BROMODOMAIN 1.
 FT DOMAIN 1612 1682 BROMODOMAIN 2.
 FT DOMAIN 1995 2068 GLN-RICH.
 FT VARIANT 575 575 P -> S.
 SQ SEQUENCE 2068 AA; 232494 MW; AD6A5ABF28B59531 CRC64;

Query Match 6.8%; Score 208; DB 1; Length 2068;
 Best Local Similarity 19.7%; Pred. No. 0.00033;
 Matches 134; Conservative 84; Mismatches 201; Indels 260; Gaps 30;

QY 2 GKKHKHKSDDKLYEE-----YKRPKLYLVKVG-NEVTELSGSG-----SGHDSL 48
 DB 1321 GSHKHERDSG---YKEVSPKPKFKLP-DLKLCGCGGVGHMRTNKKACPLYSGMSSL 1376
 QY 49 FEDK-----NDHOKHKDR-----KPK-----KRAKGEK-----71
 DB 1377 SOSNPSLADPDQSEKEMTDDDLVNDGTVTLSSKILKHGGDGRNRSSSGFT 1436
 QY 72 -QIPGEKRRKRRRVEDKKRRDRHVENAEKDLQCHAPVRLDLPPEKPLTSLAKOE 130
 DB 1437 LKVPROMCKKRRV-----GGDLHCVDLQIRNKANKANRRDPV-----1475
 QY 131 VEGTPLEALNOLMROLKRDSPAFSPVTDPIAFGYSMIINKHMFSTMKERIKNDY 190
 DB 1476 VYLSSTLEIIEHLEKSMPPVSP---FLFPVSAKVPDYRYVTKPMDLOTMRREYIQRRY 1532
 QY 191 OSIIEELKDNFKLMCTNAMTYNKPETIYKAARKLHSGKIIS--QERQSLKQSTDFPA 248
 DB 1533 TSREMFLELDKQVDSLYINGQSAVTYLAQMSESSCELLAEKDKMLREKALNPIL 1592
 QY 249 D-----LQTRKQKDGTDTSQSGEDGCGWOREDSGDAEHAFAKSPKRNK 295
 DB 1593 DDDQVALSFIFDKLHSGQKOL-----PESMPLKRP--VAK 1626
 QY 296 KDKDM-----LEDKFKSNLEREQDLDRIVKESGKILTRLVNSQCE-----339
 DB 1627 KQVKDYTYIKRPMDELITGKNIEARHYSRAEYLA-----DIELIATNCEQYNGSDT 1679
 QY 340 -----FERKRPDGTTLGLHPYDPIGEPGYCLVRLGMTTGLQSGVNTLQGFKEKRNK 395
 DB 1680 RTTKFSKTLLEYAQTO-----LIEFSEHCQGLENNIAKTO--ERAREN 1720
 QY 396 VMPVLYLNGPYSSAYAPHYSTPANISKDSDLIYSTGEDESDLPDSFIHFLATCODY 455
 DB 1721 -----APERDEAMGN---DDYMF-----DRGSRASSPD-----DY 1748
 QY 456 PTVVADSLIDVYTKGSHSTTLOEMKSLPDEBGRITLDTGKEMQITEVEPPG-----509
 DB 1749 -----IDVGHGSHASSNSIHRSMGAEGASSHTAPA-----VRKPAPPGCEYKR 1794
 QY 510 -----RLDSTODRLALAKAVTNFGVPEVTFDSEEAELIPQKIDETTRLLR 555
 DB 1795 GGRGRKORDPYEEDLQCTDDE-----DDDEEDRQ-----1826
 QY 556 ELQEAONERTLSTRPPGNMI 574
 DB 1827 EYSEEDENNAASTLDGGERI 1845

RESULT 7
 SPT7_YEAST
 ID SPT7_YEAST STANDARD; PRT; 1332 AA.
 AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Transcriptional activator SPT7.
 GN SPT7 OR YBR081C OR YBR0739.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95229044; PubMed=7713415;
 RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
 RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
 RT important for transcription in vivo."
 RL Genetics 139:523-536(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=95076715; PubMed=7985423;
 RA van der Aart O.J.M., Barthe C., Dolignon F., Aigle M., Crouzet M.,
 RA Steensma H.Y.;
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
 RT Saccharomyces cerevisiae chromosome II."
 RL Yeast 10:959-964(1994).
 RN [3]
 RP SEQUENCE OF 1-835 FROM N.A.
 RC STRAIN=5288C;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Vissers S.;
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 463-523 FROM N.A.
 RX MEDLINE=92285152; PubMed=1350857;
 RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,
 RA Dawid I.B.;
 RT "The bromodomain: a conserved sequence found in human, Drosophila and
 RT yeast proteins."
 RL Nucleic Acids Res. 20:2603-2603(1992).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
 CC OTHER GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22537; AAC37424.1; -
 CC EMBL: X76294; CAA53940.1; -
 CC EMBL: Z35950; CAA85026.1; -
 CC EMBL: M87651; AAA5087.1; -
 CC PIR: S41552; S41552.
 CC SGD: S0000285; SPT7.
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; Bromodomain1.
 DR PRINTS: PR00503; Bromodomain.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS00633; BROMODOMAIN 1; 1.
 DR PROSITE: PS50014; BROMODOMAIN 2; 1.
 KM Transcription regulation; Nuclear protein; Activator; Bromodomain.
 FT DOMAIN 458 528 BROMODOMAIN.
 SQ SEQUENCE 1332 AA; 152616 MW; 083B6362469244F CRC64;

Query Match 6.5%; Score 199.5; DB 1; Length 1332;
 Best Local Similarity 19.6%; Pred. No. 0.00038;
 Matches 141; Conservative 118; Mismatches 232; Indels 227; Gaps 31;

```

QY 20 EKPLKLVKVGNEVELSTGS-----SCHDSSLFEDKNDHKHKKRKKKGEKQI0PG 75
D 275 EERRVLNLSIKETLSKTLKNNVEIMGMNKKIYH-SFEYDEKEMIKRLKEESDCKMEK 333
QY 76 EERKRRRRVDEKDKKRDROVENEAEKDLOCHAP-----VRLDLPPE-- 118
D 334 GKRRRSNDLAAATDEODRENTNDEPTNOKLPPREGSTSDTGKRRKPSQNLDTYVNLG 393
QY 119 -----KPLTSLAKQ-----EVEEPTPLQALNOLM 144
D 394 IENSLIKHLSSIOQKSGSLGTSYELKHLIMDYRKNNKSKWTSPDRIGQELVEACEKVV 453
QY 145 ROLOR-KDPSAFSEFPYTDPAFGYSMTIKHPDPSTYKKEIKKNDIOSIELKDNFKLM 203
D 454 LELNRYTEHSTPLFNKYSKREAPYHOIIRKSMDLNTVLLKLSFOYDSKQEFVDIMLI 513
QY 204 CTNAMYKRPETIYK----- 219
D 514 WKNCILTNSDSHFLRGHAIAMOKKSLOLIMITNITIRNADLEKELEDMEKDKYELD 573
QY 220 -----AAKRLHSGMKILSQERIQ-SLKOSIDFMAD-----LOKTRKQKDG 259
D 574 EEEVAGSGRGLMGMAMLAKEKNGKSKTVKDEAPFTNDKLSVLPBEGEKEDK 633
QY 260 TDTS-----QSGEDGCGMOREREDSGDAEHAFAKSPSE--NKKKXDMLEDK 305
D 634 TASSTVTVHENVNKNEIKENG--KNEQDQVVE-ESSKTEDSSKADAKKDEGJLQDK 689
QY 306 FKSNI-----NLEREOQLDRIVKESGGLTRRLVNSOCEFRRRKPD-----GTTTL 351
D 690 TAEKKEGKGNNEEDDDDEDED-----WDSQSYLLEKDDRDDELSVWKVYTA 742
QY 352 GLHPVPDPIVEPGCYLVRIG-MTTRGLQSGVNTLQGEKEDRKKNVYPLVLYNTGPRYSY 410
D 743 KVRAEI-----CLKRTEYFKNKLNSDEAF--LKNQGRMKRFDQLFLKXEQKAL 791
QY 411 APHYDSTFAN-----ISKDSDLI-----YSTYGEES--DLPSDESIHE--FLAT 451
D 792 ESYNOKIEQNSIMKNGFTYVAKQEDDQLOPHNDHSLNGNFAFKQPDILDLDTFL-- 849
QY 452 QADPYVWADSLDLVLTFGHRSRTLOEMEMSLPEDEGHTPLDQKMEQITEVEPPGRL 511
D 850 -QEDD--ISNAIPDIYVGVNTKTLDKM-----EDASVDRMLQNG----- 886
QY 512 DSSQDRLIALKAVTNGVPEVDFDSEAEIQQ--KKLDFTTLLRELO---EAQNER 564
D 887 -INKOSRFLANK---DLGLTPKM--NONITLLOIRHICHKISILRMQSPLSAQNSR 938

```

```

CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZN-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
D EMBL: S66385; AAB28651.1; -.
D TRANSFAC: T01318; -.
D MGD: MGI:1098280; Crebbp.
D InterPro: IPR001487; Bromodomain.
D InterPro: IPR003101; KIX.
D InterPro: IPR000197; TAZ_finger.
D InterPro: IPR000433; ZnF_ZZ.
D Pfam: PF00439; bromodomain; 1.
D Pfam: PF02172; KIX; 1.
D Pfam: PF02135; zf-TAZ; 2.
D Pfam: PF00569; ZZ; 1.
D PRINTS: PR00503; BROMODOMAIN.
D SMART: SM00297; BROMO; 1.
D SMART: SM00291; ZnF_ZZ; 1.
D PROSITE: PS00633; BROMODOMAIN_1; 1.
D PROSITE: PS50014; BROMODOMAIN_2; 1.
D PROSITE: PS01357; ZF_ZZ_1; 1.
D PROSITE: PS50135; ZF_ZZ_2; 1.
D KX transcription regulation; Nuclear protein; Activator; Bromodomain;
D zinc_finger.
D FT DOMAIN 1104 1176 BROMODOMAIN.
D FT ZN_FING 1702 1745 ZZ-TYPE.
D FT DOMAIN 1062 1065 POLY-GLU.
D FT DOMAIN 1556 1563 POLY-GLU.
D FT DOMAIN 1944 1949 POLY-PRO.
D FT DOMAIN 1968 1971 POLY-GLN.
D FT DOMAIN 2082 2086 POLY-GLN.
D FT DOMAIN 2200 2216 POLY-GLN.
D FT DOMAIN 2296 2299 POLY-GLN.
D SQ SEQUENCE 2441 AA; 265474 MW; 0ABB028C3112F419 CRC64;

```

Query Match 6.4%; Score 198; DB 1; Length 2441;
 Best Local Similarity 19.9%; Pred. No. 0.0015;
 Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

```

QY 32 NEVELSTGSSGSDSLFEDKNDHKHKKRKKKGEKQIPBEGKRRRRVKED-- 88
D 971 NRVPTSTVTSATETSS--QCGPDPVPMLEKTEVQTDPAEPEPTESKGEPRSEWMEEDLO 1028
QY 89 ---KKKRDROVENEAEKDLOCHAPVRLDLPPEKPLSSIAKQEVQOT----- 134
D 1029 GSSQVKEETDTTEQKSEP-----MEVEKKRPYKVEKKEEENSNDTASQSTSPS 1079
QY 135 -----PLQALNOLMROLQKDPASAF-FSEPYTDFI--APGYSMTIKHPDSTFM 181
D 1080 QPKRKIRPEELRQALMPTLALTRQDPESLPFQAPDPOLGLIPDVFYKKNPMDLSTI 1139
QY 182 KEKIRNDYOSIEELKDNFKIMCTNAMIYKRPETIYKKAANKLHSGKILSQERIOSLK 241
D 1140 KRKLDTGQYQEPQWYVDVDRMLFNNAMLYNKRKTSRYKFKCSKL---AEVEGE-IPVVM 1194
QY 242 QSIDFMADLOTRKQKQDSTFI-----SSGSGEGCGM---QPRE 277
D 1195 QSLGCGG---RKYESPQTLCCYQKQLCTIPDAAYSYQYKHFQKGFTEIGENV 1250
QY 278 DSGDAEHAFAKSPKKNKKDMLEDFKSNNE-----RQEOQL----- 318
D 1251 TLGD-----DPGQPTTISKDQFEKK-KNDITDLPPEFVQCKBGRKMHQICVLHYDII 1302

```

QY 319 -----DRIVEGSG-----GKTRRLVNSQCFERRKRDGTTGLLHP 356
 DB 1303 WPSGFWCDLCKTKTGPRKRNKFSANRLQTRGLGNHLEDRVNFRRON-----HP 1353
 QY 357 VDPVIGEPGCLVRLGTMGTGRLOSGVNTLQGFEDKRNKVTPL-----YLYNGPYSSYAPH 413
 DB 1354 -----EAGEVFRVYVASS-----DKTVEYKPGMKSRFPVDSGMSSEFPY 1392
 QY 414 -----YDSTFANISKDSDLI-----YSTYGEDSDLPDSFSIHEFLAT 451
 DB 1393 RFRALFAPEIDGVDVCFGMHYDTALLAPHQIGCCVISTYLD-----SIHFRRR 1444
 QY 452 C-----ODPYVADSLIDVLTGCG-----HSRTLQEME 480
 DB 1445 CIRTAYVHEILIGLYEYVKRYLVYVTAHWACPPSEGDYIFHCPPDQKIPKRLQEWY 1504
 QY 481 MSLEPEGHTRLDGKEMEQITEVPPGRIDSSODRLIAKAVNPGVPVPEFSEEA 540
 DB 1505 KKLMDKAFARRLINDYKDI-----FKQANEDRLTSAELPYF-----EG 1543
 QY 541 EIFOKKIDETTRLLRELQEAONER-----LSTRPPGN 572
 DB 1544 DFWPNVLEES-----IKELQEEERKKESTASSETPEGS 1579

RESULT 9
 YC40_HUMAN STANDARD; PRT; 733 AA.
 AC 09UL10;
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein KIAA1240 (Fragment).
 GN KIAA1240.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 6:337-345(1999).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as their content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL; AB033066; BAA86554.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS0014; BROMODOMAIN_2; 1.
 KW Bromodomain; Hypothetical protein.
 FT NON_TER
 FT DOMAIN 238 320 BROMODOMAIN
 SQ SEQUENCE 733 AA; 83861 MW; 445D3D109D7F817A CRC64;

Query Match

6.3%; Score 195; DB 1; Length 733;

Best Local Similarity 20.3%; Pred. No. 0.00049;
 Matches 112; Conservative 103; Mismatches 204; Indels 134; Gaps 23;
 QY 82 RRRKVEDKKRRDRVENEAE--KDLOCHA-----PVRILDPPEKLTSSIAKOEY 131
 DB 176 QRIEDRRKFFQELLINQASAPPRKKAALCAMGLPLALPSPRROLSEBKSMEHQ 235
 QY 132 EQTPLOEA--LNQMLRQLQKDPASFSFVYDTLAPYSMAIKHPMDSFTMKERIKNN 188
 DB 236 EENTLRELRLFLRDYTKRLATDKRFNIFSKPVSDYLE-----VKEPMDLSTVITKIDKH 290
 QY 189 DVQSTIELKDNKRLKMTNMIYK-----PETIYKAKKILHSGMKL 232
 DB 291 NYLTANDPLKIDILICSNALFYNPKDPDKIIRHACPLIKPLAHLIAELDPEFNKIC 350
 QY 233 SOERISLKOSIDFEMD-----LQKT-----RKQDGTDSQSGEDGCMQRE 275
 DB 351 EELKEKRIKRLGLSVYSEQINPHSTGARKTETVEEAERKQKNPMDY-----W--- 398
 QY 276 REDSGDAEHAHAKSPSKENKKDKDMLLEDKFSNNLERPOEL-----DRIVKES 325
 DB 399 -HNSANKCAFRRYRKRRSRSSQWKGIIK-KRKYNNLKKDEEDTKFADYENHTEDEKLEEN 456
 QY 326 GGLTRRLVNSQCFERRKRDGTTGLLHPDPIVGEKCYCLVRLGTMGTG-RLOSGVNT 384
 DB 457 G----EFVSTQCHEENGEGETDLSW-----TNDSSCDI-MDLQGGRLNNGAGT 502
 QY 385 LOGFKEDKRNKTPVLYLYLYGYSSAPHYDSTFANISKDSDDLTYSTGEDSDLPDSFS 444
 DB 503 KENFASTEESSNESILVSS--SLINPQOTSR-----KETPLKNGCLNGEAS 548
 QY 445 IHEF---LATQODPYVADSLIDVLTGKHSRTLQEMEMSLPDEG---HTRTLDTG 496
 DB 549 TDSFEGIPVLEQ-----NGKLEVVSFCDSDGKSSQKILLEDQSKREKPEFSTENHG 601
 QY 497 KEMEQITEVPPGRIDSSODRLIAKAVNPGVPVPEVDSSEAEFLQKIDETTRLLRE 556
 DB 602 DDLKLE-----ALECSNNKTL-----EPGSDVEVADALDEKAGAKVKKYKRLI-- 646
 QY 557 LOEAQNERLSTRP 569
 DB 647 LEQAKTSLIELVP 659

RESULT 10
 P300_HUMAN STANDARD; PRT; 2414 AA.
 ID P300_HUMAN
 AC 009472;
 DT 15-JUL-1998 (rel. 36, Created)
 DT 15-JUL-1998 (rel. 36, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE E1A-associated protein p300.
 GN EP300 OR P300.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95011587; PubMed=7523245;
 RA Eckner R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,
 RA Lawrence J.B., Livingston D.M.;
 RT "Molecular cloning and functional analysis of the adenovirus E1A-
 RT associated 300-kD protein (p300) reveals a protein with properties of
 RT a transcriptional adaptor."
 RL Genes Dev. 8:869-884(1994).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
 CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
 CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
 CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
 CC PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.

Query Match	Similarity	6.2%	Score 190.5	DB 1	Length 2414
Best Local	Similarity	20.4%	Pred. No. 0.004		
Matches 130	Conservative	83	Mismatches 249	Indels 175	Gaps 26
38	STGSSGHDLSLEFDKNDHKKHKKRRKKRKGKQIPIGEKGRKKRRRVKEDKKRRDRDV	97			
956	STSSSTEVNSQALNAEK---QPSQGVKMEAKMEVDPPEADTQPDIDISEKVEDCKMESTET	1012			
97	ENEAEDLOCNAFVRLLDLPPEKLTSSLA---KQEEVETPLQEQALNQLMROLQRRKDS	153			
1013	E-ERSTELTETKEEDDQPTSATQSSPAGQSKKKIKFEELRIQALMPTLEALYRRDPE	1071			
154	AF-FSEPVYDFI--AEGYSMIILKHPDSTPMKEKIKKNDQSIIEELKDNFKLTNMIY	210			
1072	SLEFRPVPVQQLGIDYDPIYVKSPPDLSTIKRKLDTGYQGEPMQYVDDLTWMEFNMMWY	1131			
211	NKPETIYYAKKALLHSGMKILSOERIQSLQSIDFMA-----DLQTRKQ	256			
1132	NKRTSRYVYKYSKLT---SEVFEOE-IDPVMQSLIGVCCGGRKLEFPSTQLCCYQKQLOTIP	1186			
257	KQDTRDSQSGEDGCG-----WQEREDSGDAEAHAFKSPSKEN-KKKDKMLDEKF-K	307			
1187	RDATYYSYGNRYHFCFEKCFNEIGESYSLGDDDSQGPOTTLINKQFSRKKKNDTLDPELVE	1246			
308	SNNLEBEQGL-----DRIYKESG-----GKLTRELY	334			
1247	CTEGCGKRMHQICVLIHNEIIPAGFVCDGCLLKASARTKRNKFTSAKRLLPSTRLGFTLENRY	1306			
335	NSQCFEERRKRPDQPTTLGLLHPVDPIYGEYGCYGLYRLGMMTGTGRGQSVNVLQGGKEDKN	394			
1307	NDFLRQNHPESEGEVYRVVHASDKTYE-----YKGMKARFVDSG-ENAESEF-----	1353			

ID	ATTRX_CAEEL	STANDARD;	PRT; 1359 AA.
AC	09U7EO: 002061;		
DT	01-MAR-2002 (Rel. 4), Created)		
DT	01-MAR-2002 (Rel. 4), Last sequence update)		
DT	01-MAR-2002 (Rel. 4), Last annotation update)		
DE	Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).		
GN	XNP-1 OR B0041.7.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditidae; Rhabditioidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID-6239;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.		
RX	MEDLINE-99365296; PubMed-10433961;		
RA	Villard L., Fontes M., Embank J.D.;		
RT	"Characterization of xnp-1, a Caenorhabditis elegans gene similar to		
RT	the human XNP/ATP-X gene."		
RL	gene 236:13-19(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Fulton R., Wohlmann P.;		
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.		
CC	-1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES		
CC	GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on lists		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/isb-sib.ch).		
CC	or send an email to license@isb-sib.ch .		
CC	-----		
DR	EMBL; AF134186; AAD55361.1; -		
DR	EMBL; AF000196; AAC24256.1; -		
DR	Wormpep; B0041.7; CEI7314.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR000330; SNF2_N.		
DR	Pfam; PF00271; helicase_C; 1.		
DR	Pfam; PF00176; SNF2_N; 1.		
DR	SMART; SM00487; DEXDC; 1.		
DR	SMART; SM00490; HELIC_C; 1.		
KW	DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.		
FT	NP_BIND 436 503		
FT	SITE 636 639		
FT	DOMAIN 67 70		
FT	DOMAIN 266 272		
FT	DOMAIN 276 281		
FT	DOMAIN 372 375		
FT	DOMAIN 603 608		
FT	POLY-LYS.		


```

QY 7 KHKSDKHLIEEYVEKPLKLVKVGNEVTELTSGSSGHSLSLFDKNDHDKHDKRKRK 66
DB 1372 RHREVEYSDSLTEKQWKLKAEETLEIE-----EYVRKSSRRKRD 1416
QY 67 KKGKQIPEGEKGRKRRRVEDKRRDRVNEAEKDLQCHAPVRLDLPPEK-ITSSL 125
DB 1417 SDAGSSPTTTS---TRSDKDESKKQKRRPRAEK-----LSNPNRLTR-- 1460
QY 126 AKQEEVQETPLQELNQLMQLQKRDPSAFSPVPTDEIAGYSMIIRKHPDSTMEKI 185
DB 1461 -KMKKIYDAVIKRYKSSGRLQS---EVFIQLPSRKEL-PEYVELIRKPPDFKKIKERI 1514
QY 186 KNNQYQSTEEELAKDNFKLMTCTAMLYNKPEITYYKAAKLLHSGKILSQEIQISLKSID 245
DB 1515 RNHHYRSLNDLEKQVMLLCQNAQTFNGLSLEYDS-----IVLQSVFTSVRKIE 1565
QY 246 FMADLQKTRKQKDGDTFSQSGEDGCGWOREDSGDAAHAFKSPSKENKK--DKDMLF 303
DB 1566 -----KEDD-----SEGEES-----EHEEGEESGESSESRSYVKIKLGRKEKAQ 1606
QY 304 DKFK-----SNMLEREQQLDR 320
DB 1607 DRLKGGRRRPSRSGRAKPVYSDDSSEEQEE-DR 1639

RESULT 13
T2D1_HUMAN STANDARD; PRT; 1872 AA.
ID T2D1_HUMAN
AC P21675;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcription Initiation factor TFIID 250 kDa subunit (TAFII-250)
DE (TAFII250) (TBP-associated factor 250 kDa) (P250) (Cell cycle gene 1
DE protein).
DE TAF2A OR CCG1 OR BA2R.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=larivageal carcinoma;
RX MEDLINE=91246200; PubMed=2038334;
RA Sekiguchi T., Nohiro Y., Nakamura Y., Hisamoto N., Nishimoto T.;
RA "The human CCG1 gene, essential for progression of the G1 phase,
RA encodes a 210-kilodalton nuclear DNA-binding protein.";
RL Mol. Cell. Biol. 11:3317-3325(1991).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=89005056; PubMed=3169001;
RA Sekiguchi T., Miyata T., Nishimoto T.;
RA "Molecular cloning of the cDNA of human X chromosome gene (CCG1)
RA which complements the temperature-sensitive G1 mutants, tsb462 and
RA ts43, of the BHK cell line.";
RL EMBO J. 7:1683-1687(1988).
RN [3]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=93196704; PubMed=7680771;
RA Rupert S., Wang E.H., Tjian R.;
RA "Cloning and expression of human TAFII250: a TBP-associated factor
RA implicated in cell-cycle regulation.";
RL Nature 362:175-179(1993).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=93196705; PubMed=8450888;
RA Hatake K., Haegawa S., Takada R., Nakatani Y., Horikoshi M.,
RA Roeder R.G.;
RA "The p250 subunit of native TATA box-binding factor TFIID is the
RA cell-cycle regulatory protein CCG1.";
RL Nature 362:179-181(1993).
CC -1- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF, AS WELL AS SERVING TO

```

```

CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL
CC FOR PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES
CC DNA-BINDING ACTIVITY.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PPM: PHOSPHORYLATED BY CASEIN KINASE II IN VITRO.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -1- SIMILARITY: TO DROSOPHILA TAFII-230. SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90359; BA14374.1; -.
CC EMBL; X07024; CA30073.1; ALT_SEQ.
CC PIR; S03005; S03005.
CC PIR; A40262; A40262.
CC TRANSFAC; T02206; -.
CC MIM; 313650; -.
CC InterPro; IPR001487; Bromodomain.
CC Pfam; PF00439; bromodomain; 2.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 2.
CC PROSITE; PS00633; BROMODOMAIN_1; 2.
CC PROSITE; PS50014; BROMODOMAIN_2; 2.
CC KW Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
CC Transcription regulation; Phosphorylation.
CC FT DOMAIN 157 165 PRO-RICH.
CC FT DNA_BIND 1195 1273 HMG BOX (POTENTIAL).
CC FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT DOMAIN 1397 1467 BROMODOMAIN 1.
CC FT DOMAIN 1520 1590 BROMODOMAIN 2.
CC FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
CC SQ SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;

Query Match 5.9%; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 0.0093;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KHKSDKHLIEEYVEKPLKLVKVGNEVTELTSGSSGHSLSLFDKNDHDKHDKRKRK 65
DB 1354 KRRRGTVHCDYLNRPKRSIHRRRTDPWVTLSTLESITINDMRDLPVTPPHRPVNAKV 1413
QY 66 KKGKQIPEGEKGRKRRRVEDKRR--DRDR-----VENEAKDLQCHAPV-- 112
DB 1414 VKDYKXII---TRPMDLQTLNENVRKRLYPSREERREHLELVKNSATYNGPKSHLTGIS 1470
QY 113 ---IDLPEKPLTSSLAQOEVEQETPLQELNQLMQLQKRDPSAF----- 155
DB 1471 OSMIDLQCEK-----LKEEKDLARLEKAINPL---LDDDDQVAFSFLINDIYVQKMA 1521
QY 156 -----FSPVPTDEIAGYSMIIRKHPDSTMEKIKINNDOVQSTIELDNFKLMQTNMI 209
DB 1522 VPDSPFPHVPAKKKVPPIYKIVAMPDEITRKNSHKHTQSRSESLDDVNLILANSVK 1581
QY 210 YNKPEITYYKAAKLLHSGKILSQ--ERISQSLK-----OSIDPMADLQ 251
DB 1582 YNGPESQYTKAQELVANCYQVLTREYDEHLQLEKDICTAKAALAEALSLDPTPGP 1641
QY 252 KTRKQKDGDTDSQS---GEDGCGWOREDS--GAEAHAFKSPSKENKKDKD----- 300
DB 1642 YTPQPPDLYDNTSLSMSRDSVFPDESNMSVLDIPSAITPEKQVYQEGEDGADABEE 1701
QY 301 -----MDEKFKSNMLEREQQLDRIVYKESGKLTRLVNSQCFERRKRDGTTT 350

```



```

QY 273 QREBDSGDVAHAFKSPSKNNKKDDML-----EDKRSNNLEEOGOLDRIYESGKUL 322
Db 282 -----AOTNHRGRKRTIHPKSKDIIYESKKKPKSRKRLQDAMKFCOSVLEL----- 328
QY 330 TRRLVNSOCFEERRKPDGTTTGLLHPVDPI-VGEPGYC-LVRLGMTTGRLOSGVN----- 383
Db 329 -----MAKKHASYNYPLEPDPVPSMNLPTFYEDYVKREPMDLGTLAKKLNDMOY 376
QY 384 -TLQFKEDKRNKKTTPVLYLYL-----GPISSIAPIH-----YDSTA--NISKDSDL 428
Db 377 QTMEDEFEEDVY----LVFKNCYFNPBDGTIVNMGRLEEVFNKSWADRPNLDDYDSDE 431
QY 429 IYSTYGEDSDLPDSF---IHEFLAT 451
Db 432 .DSRTQGDYDYESYXSDDIDETILIT 457

```

Query Match	5.5%	Score 167.5	DB 1	Length 686
Best Local Similarity	21.3%	Pred. No. 0.016		
Matches 95	Conservative 76	Mismatches 180	Indels 95	Gaps 19

THIS PAGE BLANK (USP)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:44:22 ; Search time 29.85 Seconds

(without alignments)
1896.034 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073

Sequence: 1 MGKKKKKKSKHLYEEVY.....FGNMICLGPSSSEKILINK 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	506	16.5	636	2 T18845	hypothetical prote
2	252.5	8.2	715	2 T12534	hypothetical prote
3	235.5	7.7	1214	2 JC2069	zinc-finger prote
4	225.5	7.3	556	2 D96791	hypothetical prote
5	210	6.8	1586	2 S39580	HBK protein - hum
6	208	6.8	2068	2 A47371	transcription init
7	205	6.7	1572	2 S45251	SNF2alpha protei
8	199.5	6.5	1332	2 S41552	probable transcrip
9	195.5	6.5	1633	2 JC5056	polyoma 1 - chic
10	198	6.4	1359	2 T34036	hypothetical prote
11	198	6.4	1359	2 S39161	CREB-binding prote
12	192	6.2	2440	2 S39162	transcription coac
13	191.5	6.2	578	2 T40984	transcription fact
14	191	6.2	733	2 T28145	RING3 kinase - chi
15	190.5	6.2	2414	2 A54277	transcription adap
16	187.5	6.1	757	2 S68142	probable transcrip
17	187	6.1	1647	2 S45252	SNF2beta protei
18	185	6.0	1022	2 T53078	homeotic gene regu
19	183	6.0	638	2 S67605	hypothetical prote
20	183	6.0	14815	2 T48155	transcription init
21	181.5	5.9	1893	1 A40262	transcription init
22	180.5	5.9	374	2 T33328	hypothetical prote
23	179.5	5.8	754	2 A56619	female sterile hom
24	176.5	5.7	522	2 C96608	hypothetical prote
25	173.5	5.6	1490	2 S32373	DNA-binding protei
26	167.5	5.5	686	2 S55955	bromodomain protei
27	164.5	5.4	1479	2 T17401	transcription regu
28	163	5.3	1613	2 S39059	protein BRG1 - hum
29	163	5.3	1638	2 A42091	transcription acti

30	162.5	5.3	665	2 B71609	hypothetical prote
31	162.5	5.3	1877	2 T21861	hypothetical prote
32	162	5.3	1790	2 S67593	transport protein
33	161.5	5.3	729	2 S68191	trialdin - human
34	161	5.2	454	2 T37933	transcription acti
35	160	5.2	1257	2 T28937	hypothetical prote
36	159.5	5.2	1250	2 T22845	hypothetical prote
37	159.5	5.2	1994	2 D86452	protein F6N18.13
38	159	5.2	1298	2 T54367	X-linked nuclear p
39	158.5	5.2	1390	2 T34137	hypothetical prote
40	157.5	5.1	1680	2 T41628	probable transcrip
41	156.5	5.1	678	2 A54514	glutamic acid-rich
42	156.5	5.1	2139	2 T18296	myosin heavy chain
43	156.5	5.1	2526	2 T20531	hypothetical prote
44	156	5.1	1959	1 A33977	myosin heavy chain
45	156	5.1	2722	2 T20532	hypothetical prote

ALIGNMENTS

RESULT 1
T18845
hypothetical protein C01H6.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C/Accession: T18845
R/Berks, M.
submitted to the EMBL Data Library, April 1996
A/Reference number: Z19030
A/Accession: T18845
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-636 <MIL>
A/Cross-references: EMBL:Z71258; PIDN:CAA95779.1; GSPDB:GN00019; CESP:C01H6.7
A/Experimental source: clone C01H6
C/Genetics:
A/Gene: CESP:C01H6.7
A/Map position: 1
A/Intons: 20/1; 198/2; 265/3; 451/3; 489/3; 525/3
C/Superfamily: bromodomain homology
F:172-227/Domain: bromodomain homology <BRO>

Query Match	16.5%	Score 506;	DB 2;	Length 636;
Best Local Similarity	26.4%	Pred. No. 3.1e-21;		
Matches 157;	Conservative 103;	Mismatches 195;	Indels 140;	Gaps 20;
QY	53	NDHKKHDKRRKKKKKGEKQIPEEK-----GKKRR-----VKEDK---KKDRD 95		
DB	57	NNSDPEKSEDESESDDEMTTPSRKTPGCGAGGKKRRAPLTDYHLKKKILARKAARD 116		
QY	96	-RVNENKEDKQCHAPRYLDLPPREK--PLTSLAKOEVEEDTPIQALNOLMOLORLORP 152		
DB	117	AEKKEVEPEYQOEVEPEKPPPPPKAPSSS-----YLPQLMODHLRLVLEKDP 168		
QY	153	SAFSPFYDTDFIAGYSNIHKPMDFSTMEKIRKNDYQSIIEELKDNFKLCTNAMYNK 212		
DB	169	EQYAFAPVTPMADYDIILITPMDLQITRENIDGKASLPARKEDCELVSNAFYNO 228		
QY	213	PETIYYAAKKLLHSGMKLISQERIOSLKOSIDF-----MADLQKTR--KQ 256		
DB	229	PNTFYFLAAKRLSNILYVYFGEQYLRFLHSLPANKIPFIVIGIRPLAPPKERTMKR 288		
QY	257	KDGDTSQSGEDGCGWQ-----REREDSGDAEAAHAFKSPKKNKKDKDLDFKFSNN 310		
DB	289	KAVKDKGHTSD--CLQYADPKVNERLSAKLPEA-----NNPKKKKKGLGFLS-- 335		
QY	311	LERBQDLRIIVKESGKLTIRLVNSQCE--FERRKPDGTTTGLLHPVPIGPEPYC 367		
DB	336	-----EKDGTIVLVNVAAGSEDEGKLENNAPRAVITIGDIVGLEE--GTPG-- 378		
QY	368	LVRIGMTTGRLOSGVNTLQGEKDKRKNVTVPVYLNTGPGYSYAPHPDSTANISKDSD 427		

```

Db 379 --MIGMADRLTS-----OAPVNTLYGPSSAPMTDSTWMTEDTD 421
OY 428 LIYSTDGEDSLPSPFSIHEFLATCODYPYVADSLDLVLTGKSHRTLOEMESLPEDE 487
Db 422 LFLRFGDKSNMNDVSMRPFVGDCEPSEITIG-SLDDLTLGSEHKTKELE----- 473
OY 488 GHTRLDTCREKQTEVEPPGRHDSSTODRLALKAATNFVGPVEFVSEAEIF---- 543
Db 474 -----NAGEEYKE-EVDNDEKNETVLSLDDVSSISNLGTEGFLDIRQOVLVPAY 525
OY 544 -----OKLDETTRLRELQAOHERLSTRPGMNICLLG 578
Db 526 ESNINENIPEFMEVNEVNHANVQOOLNHSQKADLAHIGHRVQOPEPMIMSVQG 580

RESULT 2
112534
hypoetical protein DKFZP434B094.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 02-Sep-2000
C:Accession: 112534
R:Wambuit, R.; Heuener, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: 217524
A:Accession: 112534
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-715 <MAN>
A:Cross-references: EMBL:AL080149
A:Experimental source: adult testis; clone DKFZP434B094
C:Genetics:
A:Note: DKFZP434B094.1
C:Superfamily: bromodomain homology
F:113-168/Domain: bromodomain homology <BRO>

Query Match 8.2%; Score 252.5; DB 2; Length 715;
Best Local Similarity 26.2%; Pred. No. 7.7e-07;
Matches 88; Conservative 53; Mismatches 102; Indels 93; Gaps 11;

OY 86 KEDKKRRDRVENAEKLOCHAPVRLDLPPEKPLTSSIAKOE-----EVEQ----- 133
Db 29 QSSSQGREDEEMKAKKELKYWRRLHDLERARLLIELKRREKRRQVAYEVOYAMEL 88
OY 134 --TPLQALINOLMOLQRRKPSAFSPVPTDPIAGYSMIIRHMDSTMKERIKNDYQ 191
Db 89 RLTPLTLLRSLVLDLODKDPARIFAQPVSLKEVPDYLDHIKHPMDFATMRRLLEAGYK 148
OY 192 STEELKDNFKLMCTNAMYNNKPEITYYKAAKLLHSGMKLTSOERQSLKQSIDPMADLQ 251
Db 149 NHEFEEDDLIIDCMKYNARDTVFYRAAVRLRDGGVVLROAR----- 193
OY 252 KTRKOKDGTDSQSGEDGCGWQEREDSGDAEA---HAFKSPSKENK---KDKMLMD 304
Db 194 -----REVDSIGLEASGMLPERNAAPRRFRFSDMEDVDRLD 231
OY 305 KRKSNL---EREOLOLRI---VKESGKILTR-RIVNSOCEPERRRPDGTTTGLILH 355
Db 233 PANRAHLGLEQELRELRLDMLDITLCAMKSSGSRKRAKLKREIALLRNK-----LSQOH 285
OY 356 PVDPIVGEGYGLVRLKMTGRLQSGVNTLQGFED 391
Db 286 S-QPLPTGPG-----LGEFEED 301

RESULT 3
JC2069
zinc-finger protein, BR140 - human
N:Alternate names: bromodomain protein
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 12-Sep-1997
C:Accession: JC2069

```

```

R:Thompson, K.A.; Wang, B.; Argaves, W.S.; Giannotti, F.G.; Schranck, D.P.; Ruoslahti
Biochem. Biophys. Res. Commun. 198, 1143-1152, 1994
A:Title: BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TF
A:Reference number: JC2069; MUID:94161726
A:Accession: JC2069
A:Molecule type: mRNA
A:Residues: 1-1214 <THO>
A:Cross-references: GB:M91585
C:Comment: This is a nuclear protein with broad tissue distribution, but is especially
C:Superfamily: unassigned bromodomain proteins: bromodomain homology
C:Keywords: DNA binding; phosphoprotein; transcription regulation; zinc finger
F:653-708/Domain: bromodomain homology <BRO>
F:23,28,41,44/Binding site: zinc (Cys, Cys, His, His) #status predicted
F:120,205,462/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status
F:276,279,293,296/Binding site: zinc (Cys) #status predicted
F:301,304,317,320/Binding site: zinc (His, Cys, Cys, Cys) #status predicted
F:330,333,350,353/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:386,389,401,405/Binding site: zinc (Cys, Cys, His, Cys) #status predicted
F:410,413,444,447/Binding site: zinc (His, Cys, Cys, His) #status predicted

Query Match 7.7%; Score 235.5; DB 2; Length 1214;
Best Local Similarity 26.3%; Pred. No. 1.4e-05;
Matches 82; Conservative 58; Mismatches 129; Indels 43; Gaps 10;

OY 82 RRRVKKKKRRDRVENAE-----KDLQCHAPVRLDLPPEKPLTSSIAKOE----- 129
Db 558 RLQTHLQSGRNCDOVGRSDEKNAKELQKSMQRLHDLERARLLVELIRREKIKRE 617
OY 130 -----EVEQTPLEALNOLMOLQRRKPSAFSPVPTDPIAGYSMIIRHMDST 180
Db 618 TIVVQIAMEMOLTPRLILRLKTLLEQDKGTGNTSEVPJSEVPDYLDHIKHPMDFET 677
OY 181 MEKIKNDYOSIELEKDNFKLMCTNAMYNNKPEITYYKAAKLLHSGMKLTSOERQSL 240
Db 678 MKQNLFAVRYLNFDDFEEDFNILVSNCKLYNKKDTIFYRAAVRLRDGGAVYQARQAE 737
OY 241 KOSIDPMADLQTRKOKDGTDSQSGEDGCGWQEREDSGDAEAHAFKSPSKENKDKD 300
Db 738 KMGIDFETGMHIFHSIA-DEATHHTEDAA--EEERVLVLENGKH--PVEQOLK---- 787
OY 301 MLEDKKSNLRE---OEOLORIKESGKILTRRIVNSOCEPERRRPD--GTTTGLILH 355
Db 788 LLERLDEVNASKQSVGRSRRAKMKEMTALRLRLAH--QRETRGDPERHGPSKSLT 846
OY 356 PVDPIVGEPGYC 367
Db 847 P-----HPAAC 852

RESULT 4
D96791
hypoetical protein F15M4.12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96791
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conway, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Matzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Conn, L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-556 <STO>
A:Cross-references: GB:AE005173; NID:g65544481; PIDN:AAPI6663.1; GSPDB:GN00141
C:Genetics:

```

A:Gene: FL15M4.12
A:Map position: 1

Query Match
Best Local Similarity 23.0%; Score 225.5; DB 2; Length 556;
Matches 108; Conservative 65; Mismatches 168; Indels 129; Gaps 19;

QY 3 KKKKKHSDKLLYEEYERPLKLVKGVNTELTSGSSGHDSLEEDKN-----53
DB 12 KKKKKGRSLDLQKRALKQOQLQRRNPNEENEELSSSSRNFSNRNRNSNSD 71
QY 54 DHDHDKRRK-----KRRKGEKQIPGEKGRRRRVKEDKKRRDRVEN 99
DB 72 DDDRRDKKHLGLNSHEGRDSSNKSQGGDL--DSDARNRKT-----DSDN 120
QY 100 EAEKDLQCHAFVRLDLPPEKPLTSSLAKOEVEQTPV--QEALNQLMQLQKRPSPAFS 157
DB 121 TGEK-----ASKATDILQKSLVESTPLPKLFIIDRVQKDTYGVYS 167
QY 158 FPPVDFIAPGYSMTIKHPMDFSTMKKIKNDVOSTEB---LKDNFKLMCTNAMYKRP 213
DB 168 DPADPEELPDYELIKNPMDFTLRKKLESGAYTTLQFEASLQDVF--LICTNAMEYNSA 226
QY 214 ETIYYKAARKLLHSGMKILSGERIQSLKQSIDFMADLQKTKKOKDGTDTSGSGEDGCMQ 273
DB 227 DTYYRQKRAWLELAKK-----DFGNLRQESDGEPSVLSQOPVYK 268
QY 274 REREDSGDAEHAFAFSPSKENKKKDKMLDEKFSNNLEREQDLDRIVKESGKL----329
DB 269 RGR-----PPSGSLK-----QLEQSLIDTTSISADAAFT 301
QY 330 -----TRLVNSQCFERKRPDGTTLCLLHPVPIVGEPCVY-----RLGWTGRQ 379
DB 302 YAGSSSLSGS--YNNRK--NPPSYCFRHAETSVAVKYGMKNVDRNRPDTYNONSASLQ 356
QY 380 -SGVNTLQGFEDKRNKVTYTPV--LYLNTGPPYSVAPHYS-----TPANI 421
DB 357 DSSFTLT-----LDDMLKQLTPTGLKAEYGYARSLARKYANIGPVAWTTANV 403

RESULT 5
S39580
HBRM Protein - human
C:Species: Homo sapiens (man)
C>Date: 13-Jan-1995 #sequence_rev13 Jan-1995 #text_change 20-Sep-1999
C:Accession: S39580
R:Muchardt, C.; Yaniv, M.
EMBL J. 12, 4279-4290, 1993
A:Title: A human homologue of *Saccharomyces cerevisiae* SNF2/SW2 and *Drosophila* brm gene
A:Reference number: S39580; MUID:94038910
A:Accession: S39580
A>Status: preliminary
A:Molecule type: mRNA
A:Residue: 1-1586 <MUC>
A:Cross-references: EMBL:372889; NID:9414116; PIDN:CA51407.1; PID:9414117
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:216-249/Region: glutamine-rich
F:1423-1478/Domain: bromodomain homology <BRO>

Query Match
Best Local Similarity 6.8%; Score 210; DB 2; Length 1586;
Matches 87; Conservative 52; Mismatches 102; Indels 122; Gaps 15;

QY 54 DHDHDKR-----KRRKKRKEKQIPG-----EER-----GRRK---83
DB 1253 DMDRRREDANPRKPRIMEDELPSWIIKDDAEVERLTCEEEBEKIFGSGRQRDQVY 1312
QY 84 -----RYKEDKKR--DRDVENAEKEDLQCHAPVRLDP 116
DB 1313 SDALTEKQMLRAIEDGNLEMEEEVRLKKRRRRNVKDPKAEKVEK-----AKRRGRP 1367

QY 117 PEKPLTSS---LAKQE-----EVEQTPLOEAL-----NOLMQLQKRPSPAF 155
DB 1368 PAEKLSPPKPLTKQMAIIDTICYNKDCSCNVEKVPNSQLEIRGSSGQLS-----EVF 1423
QY 156 FSPVDFIAPGYSMTIKHPMDFSTMKKIKNDVOSTEBELKDNFKLMCTNAMYKRPET 215
DB 1424 IOLPSKREL--PEYELTRKVPDFEKIKERIRNHRYSGLDEKDVMLCHNAQTFNEGS 1482
QY 216 IYYKAARKLLHSGMKILSGERIQSLKQSIDFMADLQKTKKOKDGTDTSGSGEDGCMQ 275
DB 1483 QIYEDSLVL-----QSV--FKSAROKIAESESDEDSNEE-----EE 1519
QY 276 REDSGDAEHAFAFSPSKENKKKDKMLDEKFS-----NLEREQDLDRIVE 324
DB 1520 DEESESEAKSVKIKILNRKDKGKGRKRPNGAKPVPVSDPSDEQDERQSE 1579
QY 325 SGG 327
DB 1580 GSG 1582

RESULT 6
A47371
transcription initiation factor IID 230k chain - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 16-Feb-1994 #sequence_rev18 Nov-1994 #text_change 15-Oct-1999
C:Accession: A47371
R:Kokubo, T.; Gong, D.W.; Yamashita, S.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Genes Dev. 7, 1033-1046, 1993
A:Title: *Drosophila* 230-kD TFIID subunit, a functional homolog of the human cell cycl
A:Reference number: A47371; MUID:93279463
A:Accession: A47371
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residue: 1-2068 <KOK>
A:Cross-references: GB:561883; NID:9385550; PIDN:AAB26991.1; PID:9385551
A>Note: sequence inconsistent with nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:133002, NCBIPI:133003)
C:Keywords: transcription initiation
F:1498-1553/Domain: bromodomain homology <BRO>
F:1620-1675/Domain: bromodomain homology <BRO>

Query Match
Best Local Similarity 6.8%; Score 208; DB 2; Length 2068;
Matches 134; Conservative 84; Mismatches 201; Indels 260; Gaps 30;

QY 2 GKKKKHSDKLLYEE-----YERPLKLVKVG--NEVELSTGS-----SGHDSLT 48
DB 1321 GSHSKERDSG--YKEVSPRKKFKLP--DLKKGACAGGVGMIRNRKACPLXSGQSSLT 1376
QY 49 FEDEK-----NDHDKHDKR-----KRR-----KRRKGEK---71
DB 1377 SQSNPSLADDFEDQSEKEMTMDDDLNVNDGTQVTLSSKLLKRGHGDDKRRSGSSSGFT 1436
QY 72 -QIPGEKGRKRRRVKEDKKRRDRVENAEKDLQCHAPVRLDLPPEKPLTSSLAKOE 130
DB 1437 LKVPDRAMGKKRRV-----GGDLICDYLQRIHNTANRRDTPV-----1475
QY 131 VEQTPLOEALNQLMQLQKRPSPAFSPVDFIAPGYSMTIKHPMDFSTMKKIKNDY 190
DB 1476 VVLSSTILEITHNELRSMPPVSP--FLFPVSAKVPDYRVYTKPPDLQTMREYINQRRY 1532
QY 191 QSLTELKDNFKLMCTNAMYKRPETIYYKAARKLLHSGMKILS--QERIQSLKQSIDFMA 248
DB 1533 TSREMFLEDLKQLVNDSLNTYGPQSAVYTLAAGWFFSCCELLAEKDLMLREKAIINPL 1592
QY 249 D-----LQTRKOKDGTDTSGSGEDGCMQREDSGDAEHAFAFSPSKENK 295

```

Db 1593 DDDQVALSFIFDKLHSQIKQL-----PESWFLKP--VNK 1626
QY 296 KKDQKM-----LEDFKSNNLREBOJLDRIYKESGKLTRLVNSOE----- 339
Db 1627 KOVDYVYIKRPMDDLETTIGKNIEAHRYSRAEYLA-----DIELVINEQVNGSOT 1679
QY 340 ----ERRKPDSTTTLGLHPVDPYIGEPGYCLVRLGKMTTGLQSGVNTLQGFKEKDRNK 395
Db 1680 RYTKFSKTLFEYAQDQ-----LIERSEHCGOLENNIAKTQ---ERARBN 1720
QY 396 VMPVLVLTNGPISVAPHDSTFANISKDDSLIYSTYGEDSDLPBDFIHEFLATCOY 455
Db 1721 ----APEDDEAMGN---DDYNF---DRGSRASSPD-----DY 1748
QY 456 PYVMADSLDLVLTGKSHPTLOEMEKSLPEDGCHTPTLTGKMEQITEVEPPG----- 509
Db 1749 ----IDVEGHGCHASSNSIHRSKGAEGASSHTATA-----VRKPARPPGEVVR 1794
QY 510 ----RIDSSTODRLALKAIVNFGVPEVEVDESEAEIIFQKKLDETRLLR 555
Db 1795 GRCRPRKORDPYVEDLQCSITDE-----DDDEEDFPQ----- 1826
QY 556 ELQEAQNERLSTRPPGNMI 574
Db 1827 EVSEDENNAASITIDGERI 1845

```

```

RESULT 7
S45251
SNP2alpha protein - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C:Accession: S45251
R:Chida, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A:Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahm
A:Reference number: S45251; MUID:94268902
A:Accession: S45251
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1572 <CHR>
A:Cross-references: GB:D61655; NID:9505086; PIDN:BA05142.1; PID:41005684; PID:9987661
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1409-1464/Domain: bromodomain homology <BRO>

```

```

Query Match 6.7%; Score 205; DB 2; Length 1572;
Best local similarity 23.5%; Pred. No. 0.001;
Matches 82; Conservative 56; Mismatches 99; Indels 112; Gaps 14;

QY 54 DHDKHD---KRRKKRKGKQIQC-----EER-----GKRR----- 83
Db 1257 DMDRRREDANRPKRPRIMEDELPSMIKDAVERLTCEFEERKIFGSGQRDRVDY 1316
QY 84 ----RVKEDKKR--DRDFVNEAEKDLQCHAPVRDLP 116
Db 1317 SDALTEKQMLRALEDGNLEMEEEVRLKRRRRANDKPAKADVER-----AKRRGR 1371
QY 117 PEKPLTSLAKQEEVEQTPLOEALNOLM-RQLOQRKDP-----AFSPVYDFIAPGYS 169
Db 1372 PAEKLSPPNPK-----LTKQNNAIIDIVINYKDSGSGROLSEVFIQLPSRKEL-PEY 1422
QY 170 MIIKHPADSTMAKEKIKNNYOSEIEELKONFKLMCTNNMILYNNPPIYYAAAKLHSGM 229
Db 1423 ELIRKPVDFEKIRIRNNHRSGLDEKVMILCHNAQTFENLGSQIYEDSTVL----- 1477
QY 230 KILSOERIOGLKSGIDFMADLOKTRKQDGTTSOSGEDGCGQORERESGDAEAAHAFKS 289
Db 1478 ----QSV-FKSRQKIAKEESEDSEDSNEE-----EEDEESESSEASVYK 1519
QY 290 PSKRNKKKDKMLEDKFKS-----NNIEREQJLDRIYKESG 327
Db 1520 KIKLNKKDKGRDKGKRRKPRNGKAKPVVSDFDSEDEDERQSESGS 1568

```

```

RESULT 8
S41552
probable transcription factor SP7 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YBR0739; protein YBR081c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Jan-1994 #sequence_revision 09-Sep-1994 #text_change 20-Sep-1999
A:Accession: S41552; S45946; S40800; S45478; S54985; S59716
R:Gansheroff, L.; Dollard, C.; Tan, P.; Winston, F.
submitted to the EMBL Data Library, July 1993
A:Reference number: S41552
A:Accession: S41552
A:Molecule type: DNA
A:Residues: 1-1332 <GAM>
A:Cross-references: EMBL:L22537; NID:9349189; PIDN:AAC37424.1; PID:9349190
R:Steenma, H.Y.; van der Aart, Q.J.M.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45932
A:Accession: S45932
A:Molecule type: DNA
A:Residues: 1-1332 <STE>
A:Cross-references: EMBL:Z35950; NID:9536341; PIDN:CA085026.1; PID:9536342; MIPS:YBR0
R:Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urestarazu, A.; Vissers, S.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45893
A:Accession: S45893
A:Molecule type: DNA
A:Residues: 1-835 <AND>
A:Cross-references: EMBL:M87651; NID:9172683; PIDN:AAA55087.1; PID:9172684
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; David, I.B.
Nucleic Acids Res. 20, 2603, 1992
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast p
A:Reference number: S40800; MUID:92285152
A:Accession: S40800
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 463-523 <HAY>
A:Cross-references: EMBL:M87651; NID:9172683; PIDN:AAA55087.1; PID:9172684
R:van der Aart, Q.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.
Yeast 10, 959-964, 1994
A:Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyce
A:Reference number: S45462; MUID:95076715
A:Accession: S45478
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-624, 'LRGKKRRI', 633-1332 <VAN>
A:Cross-references: EMBL:X76294
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
R:Gansheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.
Genetics 139, 523-536, 1995
A:Title: The Saccharomyces cerevisiae SP7 gene encodes a very acidic protein importa
A:Reference number: S54985; MUID:95229044
A:Accession: S54985
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1332 <GAM>
A:Cross-references: EMBL:L22537; NID:9349189; PIDN:AAC37424.1; PID:9349190
R:van der Aart, Q.J.M.
submitted to the EMBL Data Library, August 1995
A:Reference number: S59702
A:Accession: S59716
A:Molecule type: DNA
A:Residues: 1-1332 <VAN>
A:Cross-references: EMBL:X76294; NID:974203; PIDN:CA053940.1; PID:e264674; PID:95583
A:Experimental source: strain S288c
C:Genetics:
A:Gene: SGD:SP17
A:Cross-references: SGD:S0000285; MIPS:YBR081c
A:Map position: 2R
C:Keywords: nucleus; transcription regulation
F:466-521/Domain: bromodomain homology <BRO>

```


C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T34036
 R/Fulton, R.; Wohldmann, P.
 A:Title: The EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid B0041.
 A:Reference number: 221466
 A/Accession: T34036
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1359 <FOL>
 A/Cross-references: EMBL:AF000196; PIDN:AGC24256.1; GSPDB:GN00019; CESP:B0041.7
 A:Experimental source: strain Bristol N2; clone B0041
 C/Genetics:
 A/Gene: CESP:B0041.7
 A/Map position: 1
 A/Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 6.4%; Score 198; DB 2; Length 1359;
 Best Local Similarity 21.1%; Pred. No. 0.0021;
 Matches 128; Conservative 102; Mismatches 266; Indels 112; Gaps 20;

```

OY 6 KKHSDKHLVEEYVERPLKLVKVGNEVTELTSGSSGHDLSLFEDKNDHDKRKRK 65
DB 102 KKSXSKKKVDOKKKKKSKK-----KRTSSS-----EDSDSEKOKSKK 143
OY 66 KKKGEKQIIGE--EKGRKRRYKEDKKKRD--RVENEAKDLQCHAPVRLDPEPK 120
DB 144 SKTKKQTSSESESESESESESESESESESESESESESESESESESESESESE 195
OY 121 LNSSLAKOEVQOTPLQELNQLMROKRDPAFSPVTPDIAPCYSLIKHPDFT 180
DB 196 SKKGLAKKKKSSSESESESESESESESESESESESESESESESESESESESE 254
OY 181 MEKIKNDYQSTIELKDNFKLCTNAMYINRPETIYAAKLL-----HSGMKILSQ 234
DB 255 SESSSESESESESESESESESESESESESESESESESESESESESESESESE 304
OY 235 EKIQSLKQSIDMADLOKTRKQKQDITISQSGEDGCGWQREDSGDAAHAFSPSK 294
DB 305 KKKRAVTLISDESEKDKSESESESESESESESESESESESESESESESESESE 364
OY 295 KKKDKMDLEDK--FKSNLREEOBOLDRIKESGGKLRRLVNSOCEPERKKPOGTT 351
DB 365 KKKKEPEKKKKGIIMDSKLOKETIDAEAEKER--KKRIEKKOKER----- 409
OY 352 GLLHVPDPIVGEPCYCLVRLGTTG-----RLQSGVNTLQGEKEDKRNKVTPL 403
DB 410 -----NGVLEEGEDLIE--MLTGTSSQRLKLSVLDPSSTVDESK--KPEV 460
OY 404 --YGYSSAYPH--YDSTFANISKDDSLIYSTGEDSDLPDPSIHEFLACODIPIY 458
DB 461 RLKHAHQHGIOTMDCAFESIDRLDTBSSGIIAHCMGLGTLTOYTFLLHVMHEK 520
OY 459 -MADSLDLVLTGKGHSRTLOEMKSLPEDEGHTRTLDTGKEKMEQITVEPPEGL 515
DB 521 ECKRVLVAVVPRKNIINNEKEFOKVLVNDDELDTIDVNE-----LDSYKTI 567
OY 516 ODRILATAA-----VTNGVPEVPEVDESEAEITFOKKLDETTLLRLRELQAN 567
DB 568 EDRRALAMWSSKTPSPVMIIGYDLFRILITVYDDPKKKPKRNRLLEKAKEDPK 627
OY 568 RPPGMATC 575
DB 628 PGPDMATC 635

```

RESULT 11
 S39161
 CREB-binding protein - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 20-Apr-2000
 C/Accession: S39161

R/Christia, J.C.; Kwok, R.P.S.; Lamb, N.; Hagivara, M.; Montminy, M.R.; Goodman, R.H.
 Nature 365, 855-859, 1993
 A:Title: Phosphorylated CREB binds specifically to the nuclear protein CBP.
 A:Reference number: S39161; MID:94019866
 A/Accession: S39161
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-2441 <CHR>
 A/Cross-references: GB:S66785; NID:9435854; PIDN:AAB28651.1; PID:9435855
 C/Superfamily: unassigned bromodomain proteins; bromodomain homology
 F,1112-1169/Domain: bromodomain homology. <BRO>

Query Match 6.4%; Score 198; DB 2; Length 2441;
 Best Local Similarity 19.9%; Pred. No. 0.0044;
 Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

```

OY 32 NEVTELTSGSSGHDLSLFEDKNDHDKRKRKRRKGEKQIIGEKKRRRKVED-- 88
DB 971 NRPVPTSTVSAETSS--QQPGPDVPMLEKRTVQTDAPPEPTESGEPSEMEEDLQ 1028
OY 89 ---KKRRDRVENEAEKDLQCHAPVRLDPEPKPLSLAKOEVQOT----- 134
DB 1029 GSSQVKEEDTDTTOKSEP-----MEVEKKPEVYKAEKEEENSNDTASQSTSPS 1079
OY 135 -----PLQELNQLMROKRDPAF--FSPVTPDI--APGYSMIRKHPDFTM 181
DB 1080 QPKKTIKFEPELQALMPLTEALYRQDPESLPQVDDPQLGIDPIYLVKAPMLDSTI 1139
OY 182 KEKIKNDYQSTIELKDNFKLCTNAMYINRPETIYAAKLLHSGMKILSQRISLQ 241
DB 1140 KRLIDGOYQEPWQVYDVDMFNAMLYNRKTSRYKFKCSKL-----AEVFEQ-- 1194
OY 242 QSIDFMADLOKTRKQKQDIT-----SQSGEDGCGW--QRENE 277
DB 1195 QSLGTCG--RKVPSPQTLCTCYKQLOCTIPRDAAYSYONRYHFCCKCFETI 1250
OY 278 DSGDAEHAHAFSPSKENKRRKDKMLDEKFSNNLE-----REDEQL----- 318
DB 1251 TLGD-----DPSQPTTISKQFEKK--KNPTLDPEPVDCKEGGRKHQICVLAHYDI 1302
OY 319 -----DRIVSESG-----GKILRLVNSOCEFERKRDGTTGLLHP 356
DB 1303 WPSGVCYDCLKTKGPRKRNKFSARLIQTRGNLEDRVNFLEKRN-----HP 1353
OY 357 VDPVGEPCYCLVRLGTTGRLQSGVNTLQGEKEDKRNKVTPL--YNTGYSSAYPH 413
DB 1354 -----EAGEVVRVVAAS-----DKTVEYKQMSKRFVDSGEMSSFPY 1392
OY 414 -----YDSTFANISKDDSLI-----YSTYGEDSDLPDPSIHEFLAT 451
DB 1393 RTKALFAFEIDGVDCFGMHVQDATALAPHQIOGCYVTSYD-----SIHFRRR 1444
OY 452 C-----ODPYVMAADSLDLVTGCG-----HSRLQENH 480
DB 1445 CLRTAVYHILIGYLEYAKLVYTAHIMACPPSSCEDDYIFCHHPDQIKPPKRIQENY 1504
OY 481 MSIPDEGHTRTLDTGKEKMEQITVEPPEGLDSTODRLIAKAVYNGVPEVDESE 540
DB 1505 KMLDKAPFAERILINDYKOT-----FKQANEDRLTSAKELPYF-----EG 1543
OY 541 EIFQKKLDETTLLRLRELQANER-----LSTRPPCN 572
DB 1544 DFWPNVLEES--IKELQEEERKKESTASSETPEGS 1579

```

RESULT 12
 S39162
 Transcription coactivator CREB-binding protein - human
 N/Alternate names: CBP; RSTS; Rubinstein-Taybi syndrome (RSTS) protein
 C/Species: Homo sapiens (man)
 C/Date: 07-Oct-1994 #sequence_revision 17-Nov-1995 #text_change 10-Dec-1999
 C/Accession: S39162; S60345; I58096

R:Chivla, J.C.; Kwok, R.P.S.; Lamb, N.; Hagihara, M.; Montminy, M.R.; Goodman, R.H.
Nature 365, 855-859, 1993
A:Title: phosphorylated CREB binds specifically to the nuclear protein CBP.
A:Reference number: S39161; MUID:94019866
A:Accession: S39162
A:Molecule type: mRNA
A:Residues: 1-2440 <CHR>
A:Note: differences with the mouse sequence are shown
R:Lundblad, J.R.; Kwok, R.P.S.; Laurence, M.E.; Harter, M.L.; Goodman, R.H.
Nature 374, 85-88, 1995
A:Title: Adenoviral E1A-associated protein p300 as a functional homologue of the transcr
A:Reference number: S60344; MUID:95174889
A:Accession: S60345
A:Status: preliminary
A:Molecule type: protein
R:Petrij, F.; Giles, R.H.; Dauwerse, H.G.; Saris, J.J.; Hennekam, R.C.; Masuno, M.; Tomi
Nature 376, 348-351, 1995
A:Title: Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activat
A:Reference number: 158096; MUID:95356817
A:Accession: 158096
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 352-356 <PERT>
A:Cross-References: GB:089354; NID:9188537; GB:S78936; NID:91041931
A:Note: this translation is not annotated in Genbank entry S78936, release 112.0
A:Note: this sequence with a termination mutation is from a patient with Rubinstein-Tayb
C:Genetics:
A:Gene: GDB:CREBBP; RTS: CBP; RSNS
A:Cross-References: GDB:437159; OMIM:180849; OMIM:600140
A:Map position: 16p13.3-16p13.3
A:Note: defects in this gene may result in Rubinstein-Taybi syndrome
C:Superfamily: unassigned bromodomain proteins; bromodomain homology
C:Keywords: phosphoprotein; transcription; zinc finger
F:462-661/Domain: CREB binding #status predicted <CBBA>
F:1111-1168/Domain: bromodomain homology <BRO>
F:1283-1311/Region: zinc finger CCCC motif
F:1707-1732/Region: zinc finger CCCC motif
F:78,381,745,1172,1533,2063,2354/Binding site: phosphate (Ser) (covalent) (by calmodulin
F:1771/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted

Query Match 6.2%; Score 192; DB 2; Length 2440;
Best Local Similarity 19.7%; Pred. No. 0.0097;
Matches 137; Conservative 87; Mismatches 216; Indels 254; Gaps 29;

36 ELTSSGSHSSLFEDKN--HDHKHRRKRRKKGKQIPGEGKRRKRRYKED-- 88
Db 982 ETSNQQCPVYVLEMTTEQAEDETPD-----PGSKGEPREMMEDDLOG 1028
Qy 89 -KKRRDRVENEAEKDLQCHAPVRLDLPPEKPLTSSLAKEVEQOT----- 134
Db 1029 ASQVKEETDLAEQKSEP-----MEVEDKKPEVKEVEEESSSNGTASOSTSPSQ 1079
Qy 135 -----PLQELNQLMQLORRKSAP--FSFPVYDFI--AGYSMLIKHPMDFSMK 182
Db 1080 PRKIFPEELRQALMPTLEALYRDPESLPFROPVDPQLGIDPYDIYANPMDLSTIK 1139
Qy 183 EKIKNDQSIIEELDKNFKILCTNAMYKPKETIYYNAKAKILHSGMKILISOERISLQK 242
Db 1140 RKLDTGOYQERPMQVVDVWLMFNAMLYNKRITSRYKFCSTL-----AEVFQD--IDPYMQ 1194
Qy 243 SIDFMA-----DLQTRKQKDGTDTSQSGEDGCG-----WQREKEDSDA 282
Db 1195 SLGCGCCRRKYEFPQTLCCYCKQCLTIPRDAAYSYONRYHFCEKCFTEIAGENVTLLD- 1253
Qy 283 EAHAFKSPKRNKKKKDKMDLDFKFSNNLE-----REOQL----- 318
Db 1254 -----DPSQQTITSKQFEKK--KNDPLDEPPYDCKCEGRKMHQILVLIHYDIIPSGF 1306
Qy 319 --DRIVESG-----GKLTIRLVNSQCEFFERRKPDGTTTGLHPEVDIV 361
Db 1307 VCDNCKKTKTGPRKRNKSAKRLQTRIGNLHLEEDVKNKFLRRQN-----HP----- 1352

Qy 362 GEPGYCLVRLGTMGRIGSQVNTLQGEKDKRNKVPVL---YLNTPYSSVAPH----- 413
Db 1353 -EAGEVFVRVAVSS-----DKTEVVRPGMKSRFVDSGSESPRYRTKAL 1396
Qy 414 -----YDSTFANISKDSDLI-----YSTYGEISDLPSDSITHEPLATC---- 452
Db 1397 FAFEEIDGVVCFPGNMHVDOTALIAPHQIOGCYVISTLD-----SHFFPRCLRTA 1448
Qy 453 -----QDYPYVADSLDLVLTGK-----HSRTLQEMENSLPE 485
Db 1449 VYHEILIGLYEYKAKLVYTAHTWACPPSGDDYIFCHPDDQKIPKPKRLQEMYKKMLD 1508
Qy 486 DEGHTRTLDGKMEQITFEVPEPPGRLDSSYQDRLLALKAVTNFGVPEVVDSEAEITFOK 545
Db 1509 KAFAEIRIINYKDI-----FKQANEDRLTSAKEPLPYF-----EGDFWPN 1547
Qy 546 KLDFTTLRLLEQAEQNER-----LSTPPGN 572
Db 1548 VLEBS---IKELQDEBERKKEESTASSETPEGS 1578

RESULT 13
T40984
transcription factor btf1 homolog SPC1450.02 - fission yeast (Schizosaccharomyces po
N:Alternate names: protein SPC191.13
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Dec-2000
C:Accession: T40984; T41225
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL data library, March 1999
A:Reference number: Z21962
A:Accession: T40984
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-578 <LYN1>
A:Cross-References: EMBL:AL049559; NID:94581510; PIDN:CA840169.1; PID:94581512; GSPDB
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL data library, March 1998
A:Reference number: Z21904
A:Accession: T41225
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-111 <LYN2>
A:Cross-References: EMBL:AL049644; NID:94678667; PIDN:CA841059.1; PID:94678679; GSPDB
A:Experimental source: strain 972h-; cosmid c191
C:Genetics: <C145>
A:Gene: SPDB:SPC1450.02
A:Map position: 3
C:Genetics: <C191>
A:Gene: SPDB:SPC191.13
A:Map position: 1
C:Superfamily: bromodomain homology
F:109-166/Domain: bromodomain homology <BRO1>
F:279-336/Domain: bromodomain homology <BRO2>

Query Match 6.2%; Score 191.5; DB 2; Length 578;
Best Local Similarity 21.3%; Pred. No. 0.0017;
Matches 123; Conservative 77; Mismatches 209; Indels 169; Gaps 24;

32 NEVTELSSGSHSS--LFEDKN--DHDKHRRKRRKKGKQIPGEGKRRKRRYKE 87
Db 8 NEVKAETDEIANDGSPQNLGDNNTQSSDGHNDENEESLSR--KRSSGATYGDLLQOEKE 66
Qy 88 DKRRDRVENEAEKDLQCHAPVRLDLPPEKPLTSSLAKEVEVQFPLQELNQLMQL 147
Db 67 SMPKKEPEPYKIKGS-----GMP-----PQQKCLAIIVQL 100
Qy 148 QKRDSPAFSPVYDF--IAPYSMTIKHPMDFSTMEKIKNDYOSIEELDKNFKMCT 205
Db 101 KTKNSAPFKVYVDPIKONIPDYPIIVKNPMDLGTIERKILTSYEVSVQEFIDMNLMS 160

QY 206 NAMIYKPEITIIYKAAKLLHSKMLISQRIOSLKOSIDFMADLOKTRKQDGTDSQS 265
 Db 161 NCFLYNCTESPVSQMGKAL-----QEVFERQKQLPDAPGPAAPVAKSKQKSTAP-- 213
 QY 266 GEGGCGHQREDSGDAEAHAFKSPSKENKKDKMLDEKFKSNLREOEQDLRIKES 325
 Db 214 -----PRTRRSSVSSTASVAASAPKASPAVLPE----- 245
 QY 326 GGLTRRLVNSQCEFERRRKPDGTTLLGLH-----PVDPVIGE-PGYC-LVR 370
 Db 246 -GPRRRKNNSSQMF-----CSTVLKELYKROYESAFPFQYVDVACDCEPDYVIK 298
 QY 371 LGMTHGLOSQV-----TLQGFKEKDKRNVTVLYLNTGYSSYAPHYSTRANISKD 425
 Db 299 EPMDLSTIOSKLNKNEYSTLEEFESD-----ILLM-----FNNCF----- 333
 QY 426 SDLIYSYGGSDSLPDSPFSHFEFLATCODPYWMADSLDLVLKGGHSLKLOEKMSLPE 465
 Db 334 -----TYN-----PQTPVH-----VMGRQLEVVFEKWEKPRFPDADLVKQ 371
 QY 486 DEGHTRTL-DTGKEMEOI-TEVEPPG-----RLDSSTQDRLIALKA----- 524
 Db 372 QEAETDALPDNGEERELMSEELINGAKFAVADKOISMLODTLEAKAKAKMNMRRPRR 431
 QY 525 --VTNGVPEVVDSEAE-----IFOKKIDETTRLLRE 556
 Db 432 DITREYGPITYAMQNELAERCNYLSAEOLSNVAELIRE 469

RESULT 14

T28145
 RING3 kinase - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28145
 R:Milne, S.; Kaufman, J.; Beck, S.
 submitted to the EMBL Data Library, May 1998
 A:Description: DNA sequencing and analysis of the chicken major histocompatibility comp
 A:Reference number: Z20475
 A:Accession: T28145
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-733 <MIL>
 A:Cross-references: EMBL:AL023516; NID:e1292539; PID:e1292549; PIDN:CAA18965.1
 A:Experimental source: clone GB12
 C:Genetics:
 A:Gene: RING3
 A:Map position: 16
 A:Introns: 64/3; 110/3; 158/1; 227/3; 351/3; 394/3; 479/3; 546/2; 650/1; 691/1
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 F:52-109/Domain: bromodomain homology <BR01>
 F:323-380/Domain: bromodomain homology <BR02>

Query Match 6.2%; Score 191; DB 2; Length 733;
 Best Local Similarity 28.5%; Pred. No. 0.0024;
 Matches 71; Conservative 43; Mismatches 103; Indels 32; Gaps 11;

QY 110 PYRLDLPPEKPLTSSLAQOEVEQTPLOEAL---NQLMKLOLRKPDSAF---FSEPVTD 163
 Db 275 PTK---PKKDLDPDSQHQTS-KKGLKLSBQLKYCGILLKELSKHAHAWAFYKPV-DA 329
 QY 164 IAPG---YSMIKHPMDSTAKEKTKNDYOSIEELKDNFKLMCTNMIYKPEITIIYKA 220
 Db 330 SALGHDHVEIKHPMDSTIKRKEMNRYHDAQEFADVRLMPFNCYKPNPDHVVAM 389
 QY 221 AKKLT---HSGMKILISOERIOSLKO-SIDFMADLOKTRKQDGTDSGSDGCMQRE 275
 Db 390 AKRLQDVFEFSAKMRPDQDASPPVSAPLIGALSKSSSESSDDEDED-----D 443
 QY 276 REDSGDAHAHAFKSPSKENKKDKMLDEKFKSNLREOEQDLRIKESGGLTRRLRVN 335

Db 444 EDDDESSSSSSSDSESSDSEEE-----RANRLAELOEQL-RAVHEQALALSGPVS 496
 QY 336 SQCEFERRK 344
 Db 497 KPKRRREK 505

RESULT 15

A54277
 transcription adaptor protein p300 - human
 C:Species: Homo sapiens (man)
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 10-Dec-1999
 C:Accession: A54277; S60344
 R:Ekner, R.; Ewen, M.E.; Newsome, D.; Geres, M.; Decaprio, J.A.; Lawrence, J.B.; Li
 Genes Dev. 8, 869-884, 1994
 A:Title: Molecular cloning and functional analysis of the adenovirus E1A-associated 3
 A:Reference number: A54277; MID:95011587
 A:Accession: A54277
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2414 <ECK>
 A:Cross-references: GB:U01877; NID:9495300; PIDN:AA18639.1; PID:9495301
 A:Note: in the authors' translation 941-Ser is shown after 961 and consequently, resi
 R:Lundblad, J.R.; Kwok, R.P.S.; Laurance, M.E.; Harter, M.L.; Goodman, R.H.
 Nature 374, 85-88, 1995
 A:Title: Adenoviral E1A-associated protein p300 as a functional homologue of the tran
 A:Reference number: S60344; MID:95174889
 A:Accession: S60344
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 552-660 <LUN>
 C:Genetics:
 A:Gene: GDB:EP300
 A:Cross-references: GDB:9862958; OMIM:502700
 A:Map position: 22q13.2-22q13.2
 C:Superfamily: unassigned bromodomain proteins; bromodomain homology
 C:Keywords: phosphoprotein; transcription; zinc finger
 F:1075-1132/Domain: bromodomain homology <R0>
 F:89,507,1136,1295,1497,1834,1977,2062,2320/Binding site: phosphate (Ser) (covalent)
 F:1734/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predict

Query Match 6.2%; Score 190.5; DB 2; Length 2414;
 Best Local Similarity 20.4%; Pred. No. 0.012;
 Matches 130; Conservative 83; Mismatches 249; Indels 175; Gaps 26;

QY 38 STGSSGHSLSLFEDKNDHDKHDKRRKKRKGKQJGPEKGRKRRVEDKKRRDRV 97
 Db 956 STSSTVNSQALAEK---QPSQEVKHEAKMEVDQPPADTOPBDISEKVEDECKMESTET 1012
 QY 98 ENAEKDLQCHAPVRLDLPPEKPLTSSLA---KOEVEQTPLOEALNOLMKLOLRKDS 153
 Db 1013 E-ERSTELTEIKEEDQGSTATSSPAPGOSKKTFPEELRLQALMTLEATLVNDE 1071
 QY 154 AF-FSEPVTDFT-AGYSMIKHPMDSTAKEKTKNDYOSIEELKDNFKLMCTNMIY 210
 Db 1072 SLPEKQVPDQLLGLPDYDIKSPMDLSTIKRKIDTGOYQEPWQYVDIWMFNWAMLY 1131
 QY 211 NKPETIYKAAKLLHSKMLISQRIOSLKOSIDFMA-----DLQKTRQ 256
 Db 1132 NKRTSVYTYGCKL-----SEVEQE-IDPVYOSLIGCCGRKLEFSPTLCCYCGKQLCTIP 1186
 QY 257 KDGTDTOSGEDGCG-----WQREDSGDAEAHAFKSPSKEN-KKDKMLDEKFF-K 307
 Db 1187 RDAYYSYQNRHYHCEKCFENIEOGESVSLGDDPSQPTTINNEQSKRKNDLDELPE 1246
 QY 308 SNMLREDEQL-----DLVKESE-----GKLTIRLV 334
 Db 1247 CTGGRKMQICVLHHEIIPAGFYCDGCKRSARTRENKFSAKRLDPTSLGTLENNV 1306
 QY 335 NSQCEFERRKPDGTTLLGLHPVDPIYGEPGYCLRLGTTGRTGSLGNSVLOGFEDKRN 394
 Db 1307 NDLRLQNHPEGSEYTVKRVNASHADKIVE-----VKPGKAKAFVDSG-EMAESEF----- 1353

```

QY 395 KVPVLYLNGPYSSYAPHNDSTFANISKDDSLIY-----STYGEDSDLPD----- 442
Db 1354 -----PYRKA-----LFAFEIDGVDLCFGGMHVOEYGSDCPPNORRYISY 1397
QY 443 -FSIHEFLATC---QDYPYVWADSLDVLTKGSHRTLOEMEMSLPDEGHTRTLDTGKE 498
Db 1398 LDSVHFRPKCLRTAVYHEILI-GYLEYVKKLGTY-----TGHIMACPPSEG 1443
QY 499 MEQITEVEPPGR-----LD-----SSTODRLIALKAVTN 527
Db 1444 DDYIFHCHPPDQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDIFKQATEDRLTSAKELPY 1503
QY 528 FGVPVEVFDSEAEIFQKKIDETRLRLRELQEAQNER 564
Db 1504 F-----EGDWPVPLEES---IKLEQEEER 1527

```

Search completed: July 11, 2002, 15:50:43
 Job time: 381 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:41:02 ; Search time 22.7 Seconds
(without alignments)

633,774 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073
Sequence: 1 MGKHKHKSDKHLVEYEVK...PENMICKLGPSSKCLINK 589

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5a_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3073	100.0	589	4	US-08-942-008-2
2	198	6.4	2441	1	US-08-194-468-2
3	198	6.4	2441	3	US-08-961-739-2
4	190.5	6.2	2414	1	US-08-227-536-2
5	190.5	6.2	2414	5	PCT-US95-04682-2
6	181.5	5.9	1872	1	US-08-188-582-14
7	181.5	5.9	1872	1	US-08-646-715-14
8	181.5	5.9	1893	1	US-08-188-582-11
9	181.5	5.9	1893	1	US-08-646-715-11
10	150.5	4.9	1312	2	US-08-687-080-51
11	149.5	4.9	1312	2	US-08-592-126-148
12	147.5	4.8	1588	5	PCT-US93-07261-11
13	147.5	4.8	1588	5	PCT-US93-07261-11
14	144	4.7	1886	4	US-08-938-105-3
15	142	4.6	1898	2	US-08-056-200-94
16	142	4.6	1898	2	US-08-800-644-94
17	142	4.6	3248	1	US-08-353-700-1
18	142	4.6	3248	5	PCT-US95-16216-1
19	141.5	4.6	2482	1	US-08-328-254-6
20	140	4.6	1939	4	US-09-310-187A-1
21	138.5	4.5	765	2	US-08-663-112-2
22	133.5	4.3	1161	4	US-09-327-536-2
23	132.5	4.3	1388	2	US-08-685-576-4
24	130.5	4.2	1618	1	US-07-853-913-4
25	129.5	4.2	1829	4	US-09-157-420-1
26	128.5	4.2	816	2	US-08-533-306A-6
27	128.5	4.2	816	2	US-08-742-923A-6

28	128.5	4.2	885	2	US-08-533-306A-4	Sequence 4, Appl1
29	128.5	4.2	885	2	US-08-742-923A-4	Sequence 4, Appl1
30	128.5	4.2	1112	2	US-08-714-402-2	Sequence 2, Appl1
31	128	4.2	1388	4	US-09-572-191-2	Sequence 2, Appl1
32	127	4.1	724	4	US-09-307-143-6	Sequence 6, Appl1
33	126.5	4.1	976	4	US-09-104-324B-4	Sequence 4, Appl1
34	125	4.1	1507	3	US-08-929-329-5	Sequence 5, Appl1
35	124.5	4.1	1388	2	US-08-685-576-1	Sequence 1, Appl1
36	124	4.0	1128	4	US-08-923-992A-6	Sequence 6, Appl1
37	124	4.0	1346	1	US-08-471-033-23	Sequence 23, Appl1
38	124	4.0	1346	2	US-08-471-044-23	Sequence 23, Appl1
39	124	4.0	1346	2	US-08-463-483A-23	Sequence 23, Appl1
40	124	4.0	1346	2	US-08-471-046A-23	Sequence 23, Appl1
41	124	4.0	1346	2	US-08-470-566B-23	Sequence 23, Appl1
42	124	4.0	1346	2	US-08-469-334-23	Sequence 23, Appl1
43	124	4.0	1346	3	US-09-300-529-23	Sequence 23, Appl1
44	123.5	4.0	532	1	US-08-285-440-5	Sequence 5, Appl1
45	123.5	4.0	532	1	US-08-630-349-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-942-008-2
; Sequence 2, Application US/08942008
; Patent No. 6133419
; GENERAL INFORMATION:
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: Nucleotide Sequences that Encode
; TITLE OF INVENTION: Phosphatidylinositol 3'-Kinase Associated Proteins and
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,008
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Glotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-008-2

Query Match 100.0%; Score 3073; DB 4; Length 589;

Best Local Similarity 100.0%; Pred. No. 5,1e+262; Mismatches 0; Indels 0; Gaps 0;

Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGKHKHKSDKHLVEYEVKPLKLVKVGNGVTELTSGSSGSDSLFEDKNDHDKND 60
Db 1 MGKHKHKSDKHLVEYEVKPLKLVKVGNGVTELTSGSSGSDSLFEDKNDHDKND 60

```

OY 61 RRRKKRKGKQIPGEEGRKRRRVEDKKRRDRVRNEAEKDLQCHAPVRLDLPPEKP 120
DB 61 RRRKKRKGKQIPGEEGRKRRRVEDKKRRDRVRNEAEKDLQCHAPVRLDLPPEKP 120
OY 121 LTSSIAKQEEVQOTPLQELNLMROLQKDPASAFSPYVDFIAPGSMITKHHMFST 180
DB 121 LTSSIAKQEEVQOTPLQELNLMROLQKDPASAFSPYVDFIAPGSMITKHHMFST 180
OY 181 MEKIKNNQYQISEELKNTKMTNAMIYNNKPEITYYKAARKLHSGKILSOERISL 240
DB 181 MEKIKNNQYQISEELKNTKMTNAMIYNNKPEITYYKAARKLHSGKILSOERISL 240
OY 241 KOSIDPMADLOKTRKOKGDTDSQSGEDGCGQWREREDSGDAEHAFAKSPSKENKKDKD 300
DB 241 KOSIDPMADLOKTRKOKGDTDSQSGEDGCGQWREREDSGDAEHAFAKSPSKENKKDKD 300
OY 301 MEDEPKSNLREBOEDLRYKESGKILTRLVNSQCEFEERKRDGTTGLLPVPI 360
DB 301 MEDEPKSNLREBOEDLRYKESGKILTRLVNSQCEFEERKRDGTTGLLPVPI 360
OY 361 VSEPGYCLVRLGMTGRLOSGVNTLOGFEKEDKRNKVTPLVLYNPGYSYAPHYDSTFAN 420
DB 361 VSEPGYCLVRLGMTGRLOSGVNTLOGFEKEDKRNKVTPLVLYNPGYSYAPHYDSTFAN 420
OY 421 ISKDDSDLIYSTYGGSDSPDSFISHEPLATCQDYPYVADSLDVLTKGHSRTLQEME 480
DB 421 ISKDDSDLIYSTYGGSDSPDSFISHEPLATCQDYPYVADSLDVLTKGHSRTLQEME 480
OY 481 MSLEPDEGHTRTLDTGKEMEOITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
DB 481 MSLEPDEGHTRTLDTGKEMEOITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
OY 541 EIFOKKLDETRRLRLQEOAONERLSTRPPGMICLLGPSSSEKILINK 589
DB 541 EIFOKKLDETRRLRLQEOAONERLSTRPPGMICLLGPSSSEKILINK 589

RESULT 2
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Bruggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-194-468-2

Query Match          6.4%; Score 198; DB 1; Length 2441;
Best Local Similarity 19.9%; Pred. No. 4, 1e-08;
Matches 139; Conservative 93; Mismatches 219; Indels 248; Gaps 29;

OY 32 NEVELSTGSSGSHDSLFEDKNDNDHKDRKRRKKEKQIPGEEGRKRRRVED--- 88
DB 971 NRVPPTSTVTAEMTSS--QPGPDVPMLEMTKEVQTDAAEPPTESGEPSEMEEDLQ 1028
OY 89 ---KKRRDRVRNEAEKDLQCHAPVRLDLPPEKPLTSSIAKQEEVQOT----- 134
DB 1029 GSSQVKEETDITTEQKSE-----MEVEKKEPEYKVEAKEEENSSNTASQSTSPS 1079
OY 135 -----PLQELNLMROLQKDPASAF-FSPVYVDFI--APGYSMITKHHMFSTW 181
DB 1080 QPRKRTFAPPELRQALMPTLEALYRQDPESLPFPQVDPQLGIDPYEDIVKKNPMDLSTI 1139
OY 182 MEKIKNNQYQISEELKNTKMTNAMIYNNKPEITYYKAARKLHSGKILSOERISL 241
DB 1140 KRKLDTGOYQEWQVVDVRLMFNFMALYNNKTSKVKYFCSKL---AEVEOE-IDPVM 1194
OY 242 QSIDPMADLOKTRKOKGDTDT-----SQSGEDGCGW---QREERE 277
DB 1195 QSLGVCQG---RKTEFSPQTLCCGKQGLCTIPDAAVYSQNNYHFGCKGFTLEQGENV 1250
OY 278 DSGDAEHAFAKSPSKENKKDKMDLEDKFSNNLE-----RQEOEL----- 318
DB 1251 TLGD-----DPSQPTTISKDOFEKK-KNDTLRDPFPVDCKEGKGMHQCIVLHYDI 1302
OY 319 -----DRIVKESG-----GKILTRLVNSQCEFEERKRDGTTGLLP 356
DB 1303 WPSGFCVDCNCLAKTRGRPKKEKFSAKRLQTLRLGNHLEDKRVNKLKRN-----HP 1353
OY 357 VDPYIPEPGYCLVRLGMTGRLOSGVNTLOGFEKEDKRNKVTPLV--YLVNPGYSYAPH 413
DB 1354 -----EAGEVFVRVVAAS-----DKTEVFRPGMKSRVDSGEMSESPFY 1392
OY 414 -----YDSTANISKDSDLI-----YSTYGGSDSPDSFISHEPLAT 451
DB 1393 RTKALFAEEIDGVVCFGHHVQDTLALAPHQIOGCYIISYLD-----SHFRPR 1444
OY 452 C-----QDYPYVADSLDVLTKG-----HSRTLQEME 480
DB 1445 CLRTAVYHAILGYLEYKVLVYTAHMACPPSGGDYIFCHPPODKITPKRLOEMV 1504
OY 481 MSLEPDEGHTRTLDTGKEMEOITEVEPPGRDLSSTODRLIAKAVTNGVPEVDSSEA 540
DB 1505 KKMLDKAFARLIINDYKDI-----FKQANEDRLTSAKELPY-----EG 1543
OY 541 EIFOKKLDETRRLRLQEOAONER-----LSTRPPGN 572
DB 1544 DFWPVNLEES--IKLEQEEERKKEESTAASSETPES 1579

RESULT 3
US-08-961-739-2
; Sequence 2, Application US/08961739A
; Patent No. 6063583
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
; TITLE OF INVENTION: METHODS FOR TREATING DIABETES MELLITUS
; CURRENT APPLICATION NUMBER: US/08/961,739A
; CURRENT FILING DATE: 1997-10-31
; EARLIER APPLICATION NUMBER: US 194,468
; EARLIER FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0

```



```

0Y 499 MEQIEVEPPGG-----LD-----SSTDDITLAKAVN 527
Db 1444 DDYIHCCHPPDDKIRPKRLOEWYKMKLDKAVSERIYHDYKDFIKFOKDRILTSKELPY 1503
0Y 528 FGVPVEFDFSEAELEFORKIDETETRLRLRELQEAONER 564
Db 1504 F-----EGDFWPNVLEES---IKLEQEEBER 1527

RESULT 5
PCT-US95-04682-2
; Sequence 2, Application PC/TUS9504682
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION
; TITLE OF INVENTION: FACTOR P300 AND USES OF P300
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,536
; FILING DATE: 14-April-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Holliday C. Helne, Ph.D.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DPCI-308Xq999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2414 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04682-2

Query Match 6.2%; Score 190.5; DB 5; Length 2414;
Best Local Similarity 20.4%; Pred. 1.9e-07;
Matches 130; Conservative 83; Mismatches 249; Indels 175; Gaps 26;

0Y 38 STGGSSGHSSSLFEDKNDNDHKHDKRRKKRKKKEKKQIPGEKGGKRRKRVKVEDKKKRRDRKV 97
Db 956 STSTSTEVNSQALAEK---OPSOEVRKMEAEVDDQPPADTQPEDISSEKVEEDCKMSTET 1012
0Y 98 ENEAEDKLOCHAPVRLDLPPEKPLRTSSLA---KQEEVQDTPLQALNQLMROLQRKDS 153
Db 1013 E-ERSELTETLEKEEDQSTSATQSSPARPGSKKIKFPEELRQALMTLEALYRQDE 1071
0Y 154 AF-FSPPYTDTF--APGYSMILKHPWDFSTMKKIKNNDYOSIEELKONFKLMCTNAMY 210
Db 1072 SLFQRPQVDPQLGLTIDPYDYKSPMDLSTIKRKIDLTQYQEPWQVDDIWMFNAMLY 1131
0Y 211 NKPETIYYKAARKLHSGKILISQERISLTKQSIDMA-----DLQTRQ 256
Db 1132 NKRTRVRYKYGSKL-----SEVEEQE-IDVWQSLGCGCRKLEFSPQTLCCYQKQJCTIP 1186

```

```

OY 257 KGDGDTSGSGDGC-----WQRRERDSDAAHAFKPSKEN-KKKDKMLEDKF--K 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1187 RDATYTSQNTYHNCCEKCFNEIOGESVSLDDPSQPTITNKQPSKRKNDITLDELVE 1246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 308 SNLNERDQD-----DRYKESG-----GKLTRLRY 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1247 CTBGRKMHQICVLHHEILMPAGVCDGCLKSGAPTRKKNKFSARLPSTRIGTFLERNR 1306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 335 NSQCEFERKAKDGTITGLHPVDPIYGEFGYLVLAGHTTGRHSGVNTLOGFEEDKRN 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1307 NDFLRQNHPSGEGVTARYVHASDKTYE-----YVGMKARFVSG-EMASFS----- 1353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 395 KVTPLVLYNCPESYVAPHYDSTFPANISKDSDLIY-----SYGSDSLPDS----- 442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1354 -----PYTKA-----LFAPEIDGVLCFGHNDYETGSDCPPQORRYIYSY 1397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 443 -FSIHEPLATC--QDYRYVWADSLDLVLNKGSHSPTLOEMENSLPEDBGHTRLTDGKE 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1398 LDSVHFPRKCLRTARYVHEILL-GYLEYVKKLGYT-----TGHIMACPSSG 1443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 499 MEQITEVPPER-----LD-----SSTDRLIAKATYN 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 DDYIFHCPRPQKIPKPKRLQEWYKKMLDKAVSERIVHDYKDIKQATEDRILTSAKELPY 1503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 528 FGVPVEVPDSEAEIFQKKIDETTRLRLKEQAQNER 564
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1504 F-----EGDPFNVLDES--IKELQEEER 1527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-188-582-14
: Sequence 14, Application US/08188582
: Patent No. 5534410
: GENERAL INFORMATION:
: APPLICANT: Tjian, Robert
: APPLICANT: Comali, Lucio
: APPLICANT: Dynalyst, Brian D.
: APPLICANT: Hoey, Timothy
: APPLICANT: Rupert, Siegfried
: APPLICANT: Tanese, Neoko
: APPLICANT: Wang, Edith
: TITLE OF INVENTION: WEINZIERL, Robert O.J.
: TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
: NUMBER OF INVENTION: NUCLEIC ACIDS ENCODING TARS AND METHODS OF USE
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRIGHTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/188,582
: FILING DATE: 28-JAN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman, Richard A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1872 amino acids
:

```


TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-188-582-14

Query Match 5.98; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 7.8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

6 KKHSDKHLYEEVEKPLKLVKGVNEVTELTSGSSGHSLSLEDKNDKDKRRKK 65
1354 KRRVGTTHVDYLNRRHKSIRRRTPMVTLSLESIIINDMDLNTYPTFPVNAKY 1413
66 RKKGEKOIPGEKGRKRRVKKR---DRDR-----VENEKDLQCHAPVR-- 112
1414 VKDYKII---TRPMDLQTLRENVKRLYPSREFREHLELVKNSATYNGPKHSLQIS 1470
113 ---LDLPEKPLTSSLAKEVEEOTPLQALNOLMRQLQRKPSAF----- 155
1471 QSMDLDCDEK-----LKEKEDKLARLEKAINPL---LDDDDVAFSFIIDNIYTKMA 1521
156 -----FSPYTDPLAGYSMTIKHPMDFSTMEKIKNDYQSTIEELKDNFKLCTNAMI 209
1522 VPDSWPFHHPVNAKFFVDYKVIYVPMDELITIRKNISKHKYQSRKESFLDDVNLILANSVK 1581
210 YNKPEITYYKAARKLHSGMKILSQ--ERIOSLK-----QSIDFMADLQ 251
1582 YNGPESQYTKTAQELIVNCYQTLFEYDEHLQLQEKDICTAKEALEAELESLEDPMPGPR 1641
252 KTRKQKGTDTQS---GEDGCGWQREKEDS--GDAEHAFFSKSPKSKKKDKD----- 300
1442 YTPQPPPLDYDNTSLMSRQSVQDESNNMSVLDIPSATPEKQYTOGEGDGDGLADEEE 1701
301 -----MLEDKFKSNLNEREOBOLDRIKESGKLTFRILVNSQCEFRKRPDGT 350
1702 GTVQOPQASVLYEDLMS---EGEDDERDAGSDEGDN---PFSAIQLSESGSDSDVG 1733
351 LGLHPVDPIYGEBCYCLVRLGMTTGLQSGVNTLQGFEDKRRKVPVLY----- 401
1754 SGGIRPQOPR-----LQE--NT---RMDMENESMSTYEGDGGASH 1791
402 -----LNGPYSSYAPH---YDSTFANI----- 421
1792 GLEDSNISTGSEYEPDKSNTQDTSSISGIGYEVSEEEDEEERQSGPSVLSQVHLSE 1851
422 SKDSDLIYTYGHSDDLPSD 442
1852 DEEDSEDFHSTAG--DSDLOSD 1871

RESULT 7
US-08-646-715-14
Sequence 14, Application US/08646715
Patent No. 563786
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dymally, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEO ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-646-715-14

Query Match 5.98; Score 181.5; DB 1; Length 1872;
Best Local Similarity 21.0%; Pred. No. 7.8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

6 KKHSDKHLYEEVEKPLKLVKGVNEVTELTSGSSGHSLSLEDKNDKDKRRKK 65
1354 KRRVGTTHVDYLNRRHKSIRRRTPMVTLSLESIIINDMDLNTYPTFPVNAKY 1413
66 RKKGEKOIPGEKGRKRRVKKR---DRDR-----VENEKDLQCHAPVR-- 112
1414 VKDYKII---TRPMDLQTLRENVKRLYPSREFREHLELVKNSATYNGPKHSLQIS 1470
113 ---LDLPEKPLTSSLAKEVEEOTPLQALNOLMRQLQRKPSAF----- 155
1471 QSMDLDCDEK-----LKEKEDKLARLEKAINPL---LDDDDVAFSFIIDNIYTKMA 1521
156 -----FSPYTDPLAGYSMTIKHPMDFSTMEKIKNDYQSTIEELKDNFKLCTNAMI 209
1522 VPDSWPFHHPVNAKFFVDYKVIYVPMDELITIRKNISKHKYQSRKESFLDDVNLILANSVK 1581
210 YNKPEITYYKAARKLHSGMKILSQ--ERIOSLK-----QSIDFMADLQ 251
1582 YNGPESQYTKTAQELIVNCYQTLFEYDEHLQLQEKDICTAKEALEAELESLEDPMPGPR 1641
252 KTRKQKGTDTQS---GEDGCGWQREKEDS--GDAEHAFFSKSPKSKKKDKD----- 300
1442 YTPQPPPLDYDNTSLMSRQSVQDESNNMSVLDIPSATPEKQYTOGEGDGDGLADEEE 1701
351 LGLHPVDPIYGEBCYCLVRLGMTTGLQSGVNTLQGFEDKRRKVPVLY----- 401
1754 SGGIRPQOPR-----LQE--NT---RMDMENESMSTYEGDGGASH 1791
402 -----LNGPYSSYAPH---YDSTFANI----- 421
1792 GLEDSNISTGSEYEPDKSNTQDTSSISGIGYEVSEEEDEEERQSGPSVLSQVHLSE 1851
422 SKDSDLIYTYGHSDDLPSD 442

```

DB 1852 DEDESEDFHSIAG-DSDDLSD 1871

RESULT 8
US-08-188-582-11
; Sequence 11, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comal, Lucio
; APPLICANT: Dylact, Brian D.
; APPLICANT: Hoev, Timothy
; APPLICANT: Rupert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-11

Query Match 5.9%; Score 181.5; DB 1; Length 1893;
Best Local Similarity 21.0%; Pred. No. 8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

```

```

QY 210 YKPEITVYKAAKLIHSGMKILSO--ERIOSLK-----OSIDPMADIQ 251
DB 1603 YNGPESQYTKTAQETVWVGYQTLEFDEHLTQLEKICTAKPAALAEALFESLIDPMTPGP 1662
QY 252 KTRKQKDGTDTSOS---GDDGCMOWEREDS--GDAEHAFAKSPSKENKKRKD----- 300
DB 1663 YTPQPPDLTYDTNTSLSMSKASVAFODESMNSVLDIPSAPEKQVTOEGEDGCDLADEEE 1722
QY 301 -----MLEDKFSNNLEREOQLDRIYKESGGKILRLVNSOCEFERRRKPDGTTT 350
DB 1723 GTVQOPQASVLEVDILMS--EGEDEEDAGSDEEDN-----PESAIQLSSEGSDDVYG 1774
QY 351 LGLHPVDPIVGPYGCIVRLGNTTGRLOSQVNTLQGFREKRNKVTPLY----- 401
DB 1775 SGGIRKQPRM-----LOE--NT-----RMDMENEBSMNSYEGDGEASH 1812
QY 402 -----LNYGPPSYSAFH--YDSFPANI----- 421
DB 1813 GLEDNSISTGSEEPDPKNTQDTSPSSIGYEVSEEBEDEEBEORSGPSVLSQVHUSE 1872
QY 422 SKDDSDLIYSTYGEDSDLPSPD 442
DB 1873 DEDESEDFHSIAG-DSDDLSD 1892

```

```

RESULT 9
US-08-646-715-11
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comal, Lucio
; APPLICANT: Dylact, Brian D.
; APPLICANT: Hoev, Timothy
; APPLICANT: Rupert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

MOLECULE TYPE: protein
US-08-646-715-11

Query Match 5.9%; Score 181.5; DB 1; Length 1893;
Best Local Similarity 21.0%; Pred. No. 8e-07;
Matches 118; Conservative 81; Mismatches 195; Indels 167; Gaps 21;

QY 6 KKHSDKHLVEEYKPLKLVKGVNEVTELTSGSGHDSLSFEDKNDHKRRKK 65
DB 1375 KKRREVTVHODYLNRPKSHRRRTPMVLTSLESIINDMDLNTPEFHPVAKY 1434
QY 66 RKKEKQILPGEKRRRRKVEDEKKR--DRDR-----VENEKDLQCHAPR-- 112
DB 1435 VKDYKILIT--TRPMQLQTLKENVKRLYSPREFREHLELYNSATYNGPKHSILOIS 1491
QY 113 ---LDLPEKPLTSLAKOEVEQTPQOALNOLMROLQRRKPSAF----- 155
DB 1492 QSMIDLQDEK-----LKEREDKLARLEKAINPL---LDDDDQYAFSFIIDNTVTKMMA 1542
QY 156 -----FSPVTFDIADGYSMIIKHPMDFSTMKEKIKNDYOSIEELKDNFKIMCTNAMI 209
DB 1543 VPDSMPFHHPVNNKFFVDPDYKVIYNPMDLETIRKNISKHKYQSPRESFLDDVNIILANSVK 1602
QY 210 YNKGETIYKAKAKKLLHSGCKIISO--ERIQSLK-----QSTDFMADLO 251
DB 1603 YNGESQYTTTADIEIVNVCYQTLTEYDEHLLTOLEKIDICTAKALAEALLESIDPMTPGP 1662
QY 252 KTRKQKQDGTDSOS---GEDGGCWQREERDS--GDAEVAHAFKSPSKKKKKDKD----- 300
DB 1663 YTPQRPDLDTNTSLMSRASVAPQDSNMNSVIDIPATPEKQVOTQCEGDDGLADEEE 1722
QY 301 -----MLEDKFSNNLERQEQDLRIYVESGKLTIRLVNSQCEFERRRKPDGTTT 350
DB 1723 GTVQOPQASVLYEDELMS---EGEDEDEADGSEDEGN---PFSAIQLSESGSDGVG 1774
QY 351 LGLHPVDPVIGEPGYCLTVRLGWTGRLQSGCVMTLQGFKEKDKNKVTPVLY----- 401
DB 1775 SGGRRPQOPRM-----LQF--NT---RDMENESMYSIEDGGEGASH 1812
QY 402 -----LNGPYSSYAPH---YDSTFANI----- 421
DB 1813 GLEDSNISYSYEEPPDKSNTPQTSFSSIGCYEYSEEEDEEEOQRSGPSVLSQVHLSE 1872
QY 422 SKDSDSLVSTYGEDSLPSD 442
DB 1873 DEEDSEDFHSIAG-DSDSLSD 1892

RESULT 10

US-08-687-080-51
Sequence 51, Application US/08687080
Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/08/687, 080

FILING DATE: 17-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Shultz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
US-08-687-080-51

Query Match 4.9%; Score 150.5; DB 2; Length 1312;
Best Local Similarity 17.8%; Pred. No. 0.00024;
Matches 133; Conservative 124; Mismatches 260; Indels 229; Gaps 29;

QY 2 GKHKHKKSKDLVEEYKPLKLVKGVNEVTELTSGSGHDSLSFEDKNDHKRRKK 43
DB 200 GOKVKEQOMELKYLKQKKEACEIRDQITSKEAOLTSSEKIYKVENEDLPKRLKETE 259
QY 44 HDSLEPEDKNDHDKRRKKRKKKGEKQILPGEKGR-----RRVVE 87
DB 260 HNSKTKMLKNDNEKALDSKKQKMEKDNSEL--EKKKEVFOGTDEQLNDLYNNHQPATVE 317
QY 88 DKRR-----RDRRVENEA-----EKDLQCHAPVRLDLPPEK-----PLTSLAKO 128
DB 318 KERKLVDCHELEKLNKESRLNLOEKSELLVEGRLOQADRHQEHIRARSLQSLANQ 377
QY 129 EYV---EQTPLEA---LNQLMROLQKRDPSAFSPVTDIADGYSMIIKHPMDFSTM 181
DB 378 LEIDGFERGPFSEROIKNFHKLVREKQ-----GEAKTANQIANDFAE 420
QY 182 KKKNNQYSIEELKDNF-KIMCTNAMIYKPE---TIYK-----AAKKLHSGK 230
DB 421 KETLKQKQIDELRDKTGTGLRIELKSELISKQKONELKAVKTELOLBESSDRILEDOE 480
QY 231 ILSQER-----TQSLKQSI---DPMADLOKTRKQKQDGTDSOSGEDGGCWQRE- 275
DB 481 LKKAEREILSKAKNSNVEELKMEVISLQNEKADLRLTKLQJEMEQLNHHHTTQEM 540
QY 276 -HEDSGDAHAHAFKSPSK-----NKKDKMLKLEKFR----- 307
DB 541 LTRKADKQDEQIRKIKSRSDLSLGLYPPNKKOLEMHLHSKKEINQOTRDLKLNKE 600
QY 308 -----SNLREQEQL---DRIYKESGKLTFRLVNSQCEFERRRKPD----- 346
DB 601 LASSQNNKHINNELKREQLSYEDKLFVDCGQ-----DFESDLRLKEELEKSSK 654
QY 347 -----GTTTGLLHPVDPIVEPGYC---LVRLGMTTGRLOSQVNTLOG----- 387
DB 655 QRAMLGATAV-YSQFITQITDENOSCCPVQCFQTEALEQIEVLSLQSLRLAPRLK 713
QY 388 -----FKEDRNK---VTPVLYNIGYSSYAHHYSTFANIKSDSLIYSTYGEDS 437
DB 714 STSELKKKREKRDMLGLVPRKQSTIDLKEKEIPELRLKLVNVRDIOQLKNDIEQET 773
QY 438 DLPSPFSIHEFLATQDYPYVADSLDLVLTGKHSRTLOEMENSLPDEGHTRTLDGK 497
DB 774 LGTITMPEESAKVC-----LTVDYTLNERRQOMELKQVERKIAQQAALQSIDLDR 823

OY	498 EMEOI-----TEVEPPRLSSODRIALAKAVINPEVPEVDSEAEI 54.2 ::: : ::: :
Dd	824 TVOOVNOEKOKHKLDTVSSKILKNKLIDODOEOLOHLKSTTN-----EL 87'0
OY	543 FKKLDETTRLLRELQ-EAONERIST 567 :: :
Dd	871 KSERLOISTNLRQOOLEEQTVELST 896

```

11 RESULT
12 US-08-592-126-148
13 : Sequence 148, Application US/08592126
14 : Patent No. 5821091
15 :
16 : GENERAL INFORMATION:
17 : APPLICANT: Gregory Dolgancv
18 : TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
19 : TITLE OF INVENTION: Polypeptides
20 : NUMBER OF SEQUENCES: 151
21 : CORRESPONDENCE ADDRESS:
22 : ADDRESSEE: Dehlinger & Associates
23 : STREET: 350 Cambridge Avenue, Suite 250
24 : CITY: Palo Alto
25 : STATE: CA
26 : COUNTRY: USA
27 : ZIP: 94306
28 :
29 : COMPUTER READABLE FORM:
30 : MEDIUM TYPE: Floppy disk
31 : COMPUTER: IBM PC compatible
32 : OPERATING SYSTEM: PC-DOS/MS-DOS
33 : SOFTWARE: PatentIn Release #1.0, Version #1.25
34 : CURRENT APPLICATION DATA:
35 : APPLICATION NUMBER: US/08/592,126
36 : FILING DATE:
37 : CLASSIFICATION: 435
38 :
39 : ATTORNEY/AGENT INFORMATION:
40 : NAME: Smoltz, Charles K.
41 : REGISTRATION NUMBER: 38,615
42 : REFERENCE/DOCKET NUMBER: 4600-0111
43 : TELECOMMUNICATION INFORMATION:
44 : TELEPHONE: (415) 324-0860
45 : TELEFAX: (415) 324-0960
46 :
47 : INFORMATION FOR SEQ ID NO: 148:
48 : SEQUENCE CHARACTERISTICS:
49 : LENGTH: 1312 amino acids
50 : TYPE: amino acid
51 : TOPOLOGY: linear
52 : MOLECULE TYPE: protein
53 : HYPOTHETICAL: NO
54 : ORIGINAL SOURCE:
55 : INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
56 :
57 : US-08-592-126-148

```

[illegible]

```

OY 182 KETIKNDYOSIEBLKDNF-KLMOTNMIYINRP-----TUYK-----AAKILNSGMK 230
Db 421 KETIKOKOIDEIRUKTKGIGLIELKSELISKKONELKNYKIELOLESSDRILEDOE 480
OY 231 ILISOER-----IOSLKOSI-----DEPAOLOKTRKOKDGTDSQSAGEDGCWORE- 275
Db 481 LIKARELSKAEKSNVETTKMEVISTIONEKADIDPTLRKIDEMEOJNHTTTRIOEMEM 540
OY 276 -REDSGDAEALAFSPSKE-----NKKKDDMLDEKFK----- 307
Db 541 LTKKADKDEQIRIKKSHSDELTSLLGYPPNKKLOLDEMDLHSSKELINOTRDLAKLNKE 600
OY 308 -----SNLDEROEOL-----DRIVKESGKLTTRLVNSOCFEERKRPD----- 346
Db 601 LASSQONKNHINNELKRRREDOSSYEDKLFYDCSSQ-----DPESDLRKEIEKSK 654
OY 347 -----GTTTGLHPVDPIYGEFGYC-----LYRKGMTTGRLOSSVNTLQG----- 387
Db 655 QRAMLAGATAV-YSOFTQLDENOSCCPCQRAFYOTAEIOEYISLOSKRLAPDKLK 713
OY 388 -----FKERKRNK-----VTPVLKLYNGPYSSVAPHDSFTANISKDSOLISTYGEDS 437
Db 714 STESELKKEKRDRDMGLVPMQSIIDLKEKEIPELRNKLONVNRDIORLKNDIEBOET 773
OY 438 DLPSQFSIHEFLATCQDYPYVMAOSLDVLTGKGSHRLOEMHSLPDEDSGHTTLDTGK 497
Db 774 LLGTIMPEEESAKVC-----LDDVTIMERFOMELKVERKIMQOAAKLOGLDLDLR 823
OY 498 EMEQO-----TEVEPPGRJDSSTODRLALKAALVNGFVPEVEFSEAEVI 542
Db 824 TYVOVNOKEOKEHKLDIVSSKIELNRKLIDDOEOJONHKSSTN-----EL 870
OY 543 FOKKIDETTRLLELO-BAONERLST 567
Db 871 KSEKLOISTINLORRQOLEQYVEJST 896

```

```

1      RESULT 12
2      PCT-US93-07261-11
3      ; Sequence 11, Application PC/TUS9307261
4      ; GENERAL INFORMATION:
5      ; TITLE OF INVENTION: PIEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREO
6      ; NUMBER OF SEQUENCES: 23
7      ; CORRESPONDENCE ADDRESS:
8      ; ADDRESSEE: John H. C. Blasdale
9      ; STREET: One Giralda Farms
10     ; CITY: Madison
11     ; STATE: New Jersey
12     ; COUNTRY: USA
13     ; ZIP: 07940-1000
14     ; COMPUTER READABLE FORM:
15     ; MEDIUM TYPE: Floppy disk
16     ; COMPUTER: Apple Macintosh
17     ; OPERATING SYSTEM: Macintosh 6.0.5
18     ; SOFTWARE: Microsoft Word 5.1a
19     ; CURRENT APPLICATION DATA:
20     ; APPLICATION NUMBER: PCT/US93/07261
21     ; FILING DATE: 19930805
22     ; PRIORITY APPLICATION DATA:
23     ; APPLICATION NUMBER: US 07/927,531
24     ; FILING DATE: 07-AUG-1992
25     ; ATTORNEY/AGENT INFORMATION:
26     ; NAME: Blasdale, John H. C.
27     ; REGISTRATION NUMBER: 31,895
28     ; REFERENCE/DOCKET NUMBER: DX0288K
29     ; TELECOMMUNICATION INFORMATION:
30     ; TELEPHONE: 201-822-7398
31     ; TELEFAX: 201-822-7039
32     ; INFORMATION FOR SEQ ID NO: 11:
33     ; SEQUENCE CHARACTERISTICS:
34     ; LENGTH: 1588 amino acids
35     ; TYPE: amino acid
36     ; TOPOLOGY: linear

```

MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 4.8%; Score 147.5; DB 5; Length 1563;
Best Local Similarity 19.5%; Pred. No. 0.00064;
Matches 131; Conservative 109; Mismatches 236; Indels 193; Gaps 32;

```

QY 50 EKNNDHDKRD--RRKKRKGKQIPEEKGK-----KRRVKKEDKKRRDDR 96
DB 29 OKKNOAKAKDLTKKESQSSSEKSLKEVNGEALKEKENETLKKLELNOKEKEENK 88
QY 97 VENAEKLDQCHAPYRLD--LPPEKPLTSLAKQEVQOTPLQELNOLMOLQKDPDA 154
DB 89 INDNDDEALKNGKNDKDKKIVPKRP-----SSVEKDLKEMELKEK-- 129
QY 135 FFSFVPTDFIAGYSMITKHPMDFSTMEKIKNNDYOSI--BELKDNFKLCTNAMYNK 212
DB 130 -----EFIKHLDKYEREKERNNWILRLSRDKLREIQLEKLNQLES 175
QY 213 PETIYYKAARK-----LHSGMKLISOERIOSLKQSIDFMADLOKTRKQK----DG--- 259
DB 176 INELKERRASRRPMVMVKQRMK---DEVDEWIKKYDEQAEKNGTDEEIKDKGDYEE 232
QY 260 -TDTSGS-----EDGCGMQREREDSGDAEHAFAKSP----- 290
DB 233 IYETFTYGMRENALGELDEYERERKRYYLKEDEGDLKVEEKLLETGTFREKFPPT 292
QY 291 -----SKENKKKDKMDLEDKFSNNLEREOQLDRI--VESGKLTTRLVNSOCFE--ERR 343
DB 293 RILVRRKRNKEQKKLKEDEKKEKLLAEBPDDEKIKLSDDKYVYVNNKKSFPDKFR 352
QY 344 KPDGTTT---LGLHPDPYVGEPCYVLRGWTGRLQSGVN-----TLQGFKKDKR 393
DB 353 APDKKRTMFYRLSELFPYVPRKDN-----ELAVGDSMSKVGKKLKSTFNFPR-RR 405
QY 394 NKVPYVLYLNGPYSSYAPHYSTFANISKD-----SDLI----- 429
DB 406 NKLKRRKQELHKFKKNNKKYQKLLEREKRNPGCEPLNPEIHVIRPSDLMKGENKSA 465
QY 430 -----YSTVG-----EDSDLPDSFI--HEFLATQDYP--YVADSLLD-----VLT 468
DB 466 GHPKRYOPTKGLKYEESHVSKDYQLEHEPPTKLPEYKGVHSREYQLDHEPPTKPEYE 525
QY 469 KGSRSRLQ---EMEMSLPE-DEGHTR---TLD-----TGKEMQITVEY-----EPP 508
DB 526 KGVHSREYQLDNEVRDELPEYKGVHSREYQLDNEGSTLKEVDQ--TELAKGKDTNNKPH 584
QY 509 GRDSSQDRLIALKAVTNGVPEVDFSEAEITFOKKLDETRLLREIQE-----AON 562
DB 585 ESVDYDQTELAKGKDTN--KPHESVDYDQSELAKGKDTNNKPHESVDYDQTELAKG 642
QY 563 ERLSTRPPGNM 573
DB 643 KEVTNNKHEML 653

```

RESULT 13
PCT-US93-07261-16
Sequence 16, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
STREET: One Giralda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ. ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
STRAIN: Malayan Camp
PCT-US93-07261-16

Query Match 4.8%; Score 147.5; DB 5; Length 1663;
Best Local Similarity 19.5%; Pred. No. 0.00064;
Matches 131; Conservative 109; Mismatches 236; Indels 193; Gaps 32;

```

QY 50 EKNNDHDKRD--RRKKRKGKQIPEEKGK-----KRRVKKEDKKRRDDR 96
DB 29 OKKNOAKAKDLTKKESQSSSEKSLKEVNGEALKEKENETLKKLELNOKEKEENK 88
QY 97 VENAEKLDQCHAPYRLD--LPPEKPLTSLAKQEVQOTPLQELNOLMOLQKDPDA 154
DB 89 INDNDDEALKNGKNDKDKKIVPKRP-----ESVEKDLKEMELKEK-- 129
QY 155 FFSFVPTDFIAGYSMITKHPMDFSTMEKIKNNDYOSI--BELKDNFKLCTNAMYNK 212
DB 130 -----EFIKHLDKYEREKERNNWILRLSRDKLREIQLEKLNQLES 175
QY 213 PETIYYKAARK-----LHSGMKLISOERIOSLKQSIDFMADLOKTRKQK----DG--- 259
DB 176 INELKERRASRRPMVMVKQRMK---DEVDEWIKKYDEQAEKNGTDEEIKDKGDYEE 232
QY 260 -TDTSGS-----EDGCGMQREREDSGDAEHAFAKSP----- 290
DB 233 IYETFTYGMRENALGELDEYERERKRYYLKEDEGDLKVEEKLLETGTFREKFPPT 292
QY 291 -----SKENKKKDKMDLEDKFSNNLEREOQLDRI--VESGKLTTRLVNSOCFE--ERR 343
DB 293 RILVRRKRNKEQKKLKEDEKKEKLLAEBPDDEKIKLSDDKYVYVNNKKSFPDKFR 352
QY 344 KPDGTTT---LGLHPDPYVGEPCYVLRGWTGRLQSGVN-----TLQGFKKDKR 393
DB 353 APDKKRTMFYRLSELFPYVPRKDN-----ELAVGDSMSKVGKKLKSTFNFPR-RR 405
QY 394 NKVPYVLYLNGPYSSYAPHYSTFANISKD-----SDLI----- 429
DB 406 NKLKRRKQELHKFKKNNKKYQKLLEREKRNPGCEPLNPEIHVIRPSDLMKGENKSA 465
QY 430 -----YSTVG-----EDSDLPDSFI--HEFLATQDYP--YVADSLLD-----VLT 468
DB 466 GHPKRYOPTKGLKYEESHVSKDYQLEHEPPTKLPEYKGVHSREYQLDHEPPTKPEYE 525
QY 469 KGSRSRLQ---EMEMSLPE-DEGHTR---TLD-----TGKEMQITVEY-----EPP 508
DB 526 KGVHSREYQLDNEVRDELPEYKGVHSREYQLDNEGSTLKEVDQ--TELAKGKDTNNKPH 584
QY 509 GRDSSQDRLIALKAVTNGVPEVDFSEAEITFOKKLDETRLLREIQE-----AON 562

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 15:32:31 ; Search time 55.8 Seconds

(without alignments)
1172.446 Million cell updates/sec

Title: US-09-687-230-2

Perfect score: 3073
Sequence: 1 MCKKKKKKKSKHLYEYVE.....PENNICLLGPSSKCLINK 589

Scoring table:

BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: A.Geneseq_032802:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3073	100.0	589	19	AAW37947
2	2995	97.5	667	22	AAO13504
3	2972.5	96.7	718	22	AAU16226
4	1710.5	55.7	405	22	AAW37112
5	1413.5	46.0	351	22	AAW3765
6	921.5	30.0	597	22	AAW38835
7	920.5	30.0	605	22	AAW40621
8	798.5	26.0	199	22	AAU16213
9	699.5	22.8	861	22	ABW63028
10	684	22.3	501	22	AAW95421
11	476	13.5	280	21	AAW41780

12	391	12.7	233	22	AAW95881	Human protein sequ
13	258.5	8.4	1058	22	AAW39231	Human polypeptide
14	258.5	8.4	1109	22	AAW41017	Human polypeptide
15	258.5	8.4	1189	22	AAW41018	Human polypeptide
16	258.5	8.4	1189	22	AAW39232	Human polypeptide
17	252.5	8.2	414	22	AAW63767	Human prostate can
18	252.5	8.2	715	22	ABG12634	Novel human diagno
19	235.5	7.7	707	22	ABG21300	Novel human diagno
20	235.5	7.7	1173	22	ABG21301	Novel human diagno
21	229.5	7.5	2543	22	ABG21295	Novel human diagno
22	221	7.2	616	21	ABG28052	Novel human diagno
23	218	7.1	1430	22	ABW58602	Arabidopsis thalia
24	215	7.0	221	21	AAW43501	Human cancer assoc
25	212	6.9	245	22	AAW63828	Human prostate can
26	210.5	6.8	2442	22	ABW44555	Human wound healin
27	208	6.8	2065	22	ABW63556	Human wound healin
28	208	6.8	2065	22	ABW63556	Human wound healin
29	198	6.4	2441	16	AAW70054	Human wound healin
30	198	6.4	2441	19	AAW40058	Human wound healin
31	198	6.4	2441	22	AAW44252	Human wound healin
32	192.5	6.3	1649	21	AAW44555	Human wound healin
33	191	6.2	1650	21	AAW27557	Human wound healin
34	191	6.2	1678	21	AAW27554	Human wound healin
35	190.5	6.2	2414	16	AAW84882	Human wound healin
36	190.5	6.2	2414	19	AAW40057	Human wound healin
37	189.5	6.2	1679	21	AAW27552	Human wound healin
38	188.5	6.1	1646	21	AAW27553	Human wound healin
39	188.5	6.1	1681	21	AAW27558	Human wound healin
40	187	6.1	1647	21	AAW27549	Human wound healin
41	187	6.1	1682	21	AAW27556	Human wound healin
42	187	6.1	1924	22	ABG06008	Human wound healin
43	186.5	6.1	1886	22	ABG06009	Human wound healin
44	183	6.0	580	22	AAU16207	Human wound healin
45	183	6.0	2234	22	ABG00985	Human wound healin

ALIGNMENTS

RESULT 1	AAW37947	standard: Protein: 589 AA.
ID	AAW37947	standard: Protein: 589 AA.
XX	AAW37947	
AC	AAW37947	
XX	AAW37947	
DT	AAW37947	
XX	AAW37947	
DE	AAW37947	
XX	AAW37947	
KW	Phosphatidylinositol-3' kinase associated protein.	
KW	human; signal transduction; cell growth; cancer; restenosis;	
KW	therapy; diagnosis.	
XX	Phosphatidylinositol-3' kinase associated protein.	
OS	Homo sapiens.	
XX	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	Domain	151..313
FT	Region	/note="bromodomain"
FT	Region	516..589
FT	Region	/note="PI3K p85 binding region"
XX	W09820126-A1.	
XX	W09820126-A1.	
XX	14-MAY-1998.	
XX	14-MAY-1998.	
XX	01-OCT-1997;	97WO-US15845.
XX	01-OCT-1997;	97WO-US15845.
XX	01-NOV-1996;	96US-0030103.
XX	01-NOV-1996;	96US-0030103.
XX	(ONVX-) ONVX PHARM INC.	
XX	(ONVX-) ONVX PHARM INC.	
XX	Braselmann S;	
XX	Braselmann S;	

DR WPI: 1998-286942/25.
DR N-PSDB: AAV29267.

PT New isolated phosphatidylinositol-3/kinase associated protein -
PT used to develop products for diagnosis and treatment of cell growth
PT disorders such as restenosis or cancer

PS Claim 10: Page 40-41: 52pp: English.

XX This polypeptide comprises human phosphatidylinositol-3' kinase
CC (PI3K) associated protein (PI3AP), a protein that binds to the
CC intermediate SH2 domain on the p85 regulatory subunit of PI3K, and
CC which exhibits a bromodomain. Its amino acid sequence was deduced
CC from a cDNA clone (see AAV29267) obtained from an HeLa library using
CC a yeast two-hybrid assay with PI3K p85 as bait. The invention
CC provides vectors containing nucleic acid sequences that encode
CC PI3AP or its fragments, host cells, methods for the expression of
CC PI3AP, and methods for using the products for the diagnosis and
CC treatment of cell growth disorders such as restenosis or cancer.
CC Also described is an assay for identifying agonists and antagonists
CC of PI3K regulation. These include mutant PI3APs that compete with
CC native PI3APs for binding to PI3K, antibodies, and nucleotide
CC sequences that can be used to inhibit or enhance PI3AP gene
CC expression. Transgenic and knock-out animals are also described.

SO Sequence 589 AA;

Query Match 100.0%; Score 3073; DB 19; Length 589;
Best Local Similarity 100.0%; Pred. No. 6.7e-248;
Matches 589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGKRRKRRKSKHLVEEYVERKPLKLVKVGNEVTELTSGSSGHSLSLFEDKNDHKDKKD 60
DB 1 mgkrrkrrkskhlveeyvekrplklvkvgnvteletsgssghsllsfedkndhdkkd 60
OY 61 RRRKRRKGEKQIPGEKGRKRRVKKEDKKRRDRVNEAEKDLQCHAPVRDLDPPEKP 120
DB 61 rrrkrrkgeqkipgeekgrkrrvkkedkrrdrvneaeekdlqchavprldlpppek 120
OY 121 LTSSIAKOEVEQRPLOALNOLMRQLOKRPDSAFSPVPDFTAPGSMIIKHPMDFST 180
DB 121 ltssiakoeveeqrploalnolmrqlorkrpdsafspvpdftapgsmiikhpmdfst 180
OY 121 ltssiakoeveeqrploalnolmrqlorkrpdsafsfipvdftlapygmlikhpmdfst 180
DB 121 ltssiakoeveeqrploalnolmrqlorkrpdsafsfipvdftlapygmlikhpmdfst 180
OY 181 MKEKIKNDYOSIEELKDNFKLMCTNAMIYNNPETYKAAKRLHSGMKLISGERRISL 240
DB 181 mkekikndyosieelkdnfkimctnamlynnpetiykaakrlhsgmkllisgeris 240
OY 241 KQSIDFMADLQTRKQKQGTDTSSGSEGGCQWRERDSGAEAHAFKSPSKENKKDKD 300
DB 241 kqsidfmadlqtrkqkqgtdtssgsedgqwcwqretdsgdaeahafkspskenkdkd 300
OY 301 MEDFKSNLNEREOQLDRIVKESGKLTRLVNSOCEFERRRKPDGTTTGLLHPVPI 360
DB 301 mledfksnlneredqldrivkesgklttrlvnsceferrrkpdgtttgllhpvp 360
OY 361 VGEPCYCLVRLQMTTGRLOSGVNTLQGRKEDKRNKVTPLYLYNPGYSSAPHTYSTAN 420
DB 361 vgepcyclvrlqmttgrlqsgvntlqgkredkrnkvtplylyngpyssaphtystan 420
OY 421 ISKSDSDILSYTGSDSDLPDSFSTHEFLATQDYPYVADSLDYLVTGSGSRTLQEME 480
DB 421 isksdssdillystgdsdldpdsfstheflatqdyvyvadsldylvtgsgsrtlqeme 480
OY 481 MSLEPDEGHTRTLDTGKENEQITVEVPPGRLDSTQDRLIALKAVTNGVPEVDESEBA 540
DB 481 mslpedeghtrtltdtgkeneqitvevppgrldstqodrlialkavtngvpevde 540
OY 541 ELFPQKRLDETTRLRLQEAQNERLSTRPPGNWICLLGSSSRKCLILNK 589
DB 541 elfpqkrldettrlrlqeaqnerlstrppgnwiclsgsssrkclilnk 589

```

RESULT 2

ID AA013504 standard; protein; 667 AA.

AA013504;

06-NOV-2001 (first entry)

Human polypeptide SEQ ID NO 27396.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001MO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB: AAI93435.

PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -

XX Claim 20: SEQ ID NO 27396; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 667 AA;

Query Match 97.5%; Score 2995; DB 22; Length 667;
Best Local Similarity 99.5%; Pred. No. 2.6e-241;
Matches 575; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 3 KKHKKKSDKHLVEEYVERKPLKLVKVGNEVTELTSGSSGHSLSLFEDKNDHKDKRK 62
DB 3 kkhkkksdkhlveeyvekrplklvkvgnvteletsgssghsllsfedkndhdkrk 62
OY 18 KKHKKKSDKHLVEEYVERKPLKLVKVGNEVTELTSGSSGHSLSLFEDKNDHKDKRK 77
DB 18 kkhkkksdkhlveeyvekrplklvkvgnvteletsgssghsllsfedkndhdkrk 77
OY 63 RRRKRRKGEKQIPGEKGRKRRVKKEDKKRRDRVNEAEKDLQCHAPVRDLDPPEKP 122
DB 63 rrrkrrkgeqkipgeekgrkrrvkkedkrrdrvneaeekdlqchavprldlpppek 122
OY 78 RKKRRKGEKQIPGEKGRKRRVKKEDKKRRDRVNEAEKDLQCHAPVRDLDPPEKP 137
DB 78 rkkrrkgeqkipgeekgrkrrvkkedkrrdrvneaeekdlqchavprldlpppek 137
OY 123 SSLAKOEVEQRPLOALNOLMRQLOKRPDSAFSPVPDFTAPGSMIIKHPMDFSTMK 182
DB 123 sslakoeveeqrploalnolmrqlorkrpdsafspvpdftapgsmiikhpmdfstmk 182
OY 138 SSLAKOEVEQRPLOALNOLMRQLOKRPDSAFSPVPDFTAPGSMIIKHPMDFSTMK 197
DB 138 sslakoeveeqrploalnolmrqlorkrpdsafsfipvdftlapygmlikhpmdfstmk 197

```

QY 183 EKIKNDYOSTEELKDNFKLMCTNAMIYKRPETIYYKAACKLHSGKIIISQERISLQ 242
Db 198 EKlknndygsleelkdnfkImctnamIynkpetIlykaakkllhsnkmllsgerIqslkq 257
QY 243 SIDFMADLOKTRKQKDGTDTSQSGEDGCGWQREKREDSGDAEAHAFKSPKRNKKKDKML 302
Db 258 sIdfmaIdqtrkqkdgtIdcsqsgedgscwqredsgdeahafkspkrenkkkdkdml 317
QY 303 EDKRSNNLEBEOQLDRIVKESGKILTRRLVNSQCFEERRKPDGTTTGLHPVDPYV 362
Db 318 edkRsnmleregeqlrIvKesgkIltrrlvnsqceferRkpqdtllglhpvdplvg 377
QY 363 EPGCYCLRIAGTWRLOSGYNTLOGFEKEDKRNKVTPLVLYNNGYSYAAHYSTFANIS 422
Db 378 epygcylrlagtmcgrIsgyvtllqgfekdkrnkvtplylnyppysyaphystfianIs 437
QY 423 KDDSDLIYSTYGEDSDLPSPFSIHFLATCODYPYVADLLVLITKGHSRTLOEWEMS 482
Db 438 kddsdIlystygEdsdIpsfIsIheflatcdqpyvmadllvlItkghsrtlgemems 497
QY 483 LPDEGHTRLTDGKMEQITTEVPEPRLDSTODRLIAKAVTNFEGVPEVDFSEAEI 542
Db 498 lPdeghtrltIdakemeqIttevpprldstqdrlIaIkavtnfgvpvdfseaeI 557
QY 543 FOKKIDETTRLRLRLOPAONERLSTRPPGMICLIGPS 580
Db 558 fOkkIdettrIrlreIgeaonerIstrppgmIclIgps 595

RESULT 3
AAU16626
ID AAU16626 standard; Protein; 718 AA.
XX
AC AAU16626:
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1579.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytosolic; cadiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN MO200155322-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01341.
XX
PF 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225478.
PR 14-AUG-2000; 2000US-0225479.
PR 14-AUG-2000; 2000US-0225480.
PR 14-AUG-2000; 2000US-0225481.
PR 14-AUG-2000; 2000US-0225482.
PR 14-AUG-2000; 2000US-0225483.
PR 14-AUG-2000; 2000US-0225484.
PR 14-AUG-2000; 2000US-0225485.
PR 14-AUG-2000; 2000US-0225486.
PR 14-AUG-2000; 2000US-0225487.
PR 14-AUG-2000; 2000US-0225488.
PR 14-AUG-2000; 2000US-0225489.
PR 14-AUG-2000; 2000US-0225490.
PR 14-AUG-2000; 2000US-0225491.
PR 14-AUG-2000; 2000US-0225492.
PR 14-AUG-2000; 2000US-0225493.
PR 14-AUG-2000; 2000US-0225494.
PR 14-AUG-2000; 2000US-0225495.
PR 14-AUG-2000; 2000US-0225496.
PR 14-AUG-2000; 2000US-0225497.
PR 14-AUG-2000; 2000US-0225498.
PR 14-AUG-2000; 2000US-0225499.
PR 14-AUG-2000; 2000US-0225500.
PR 14-AUG-2000; 2000US-0225501.
PR 14-AUG-2000; 2000US-0225502.
PR 14-AUG-2000; 2000US-0225503.
PR 14-AUG-2000; 2000US-0225504.
PR 14-AUG-2000; 2000US-0225505.
PR 14-AUG-2000; 2000US-0225506.
PR 14-AUG-2000; 2000US-0225507.
PR 14-AUG-2000; 2000US-0225508.
PR 14-AUG-2000; 2000US-0225509.
PR 14-AUG-2000; 2000US-0225510.
PR 14-AUG-2000; 2000US-0225511.
PR 14-AUG-2000; 2000US-0225512.
PR 14-AUG-2000; 2000US-0225513.
PR 14-AUG-2000; 2000US-0225514.
PR 14-AUG-2000; 2000US-0225515.
PR 14-AUG-2000; 2000US-0225516.
PR 14-AUG-2000; 2000US-0225517.
PR 14-AUG-2000; 2000US-0225518.
PR 14-AUG-2000; 2000US-0225519.
PR 14-AUG-2000; 2000US-0225520.
PR 14-AUG-2000; 2000US-0225521.
PR 14-AUG-2000; 2000US-0225522.
PR 14-AUG-2000; 2000US-0225523.
PR 14-AUG-2000; 2000US-0225524.
PR 14-AUG-2000; 2000US-0225525.
PR 14-AUG-2000; 2000US-0225526.
PR 14-AUG-2000; 2000US-0225527.
PR 14-AUG-2000; 2000US-0225528.
PR 14-AUG-2000; 2000US-0225529.
PR 14-AUG-2000; 2000US-0225530.
PR 14-AUG-2000; 2000US-0225531.
PR 14-AUG-2000; 2000US-0225532.
PR 14-AUG-2000; 2000US-0225533.
PR 14-AUG-2000; 2000US-0225534.
PR 14-AUG-2000; 2000US-0225535.
PR 14-AUG-2000; 2000US-0225536.
PR 14-AUG-2000; 2000US-0225537.
PR 14-AUG-2000; 2000US-0225538.
PR 14-AUG-2000; 2000US-0225539.
PR 14-AUG-2000; 2000US-0225540.
PR 14-AUG-2000; 2000US-0225541.
PR 14-AUG-2000; 2000US-0225542.
PR 14-AUG-2000; 2000US-0225543.
PR 14-AUG-2000; 2000US-0225544.
PR 14-AUG-2000; 2000US-0225545.
PR 14-AUG-2000; 2000US-0225546.
PR 14-AUG-2000; 2000US-0225547.
PR 14-AUG-2000; 2000US-0225548.
PR 14-AUG-2000; 2000US-0225549.
PR 14-AUG-2000; 2000US-0225550.
PR 14-AUG-2000; 2000US-0225551.
PR 14-AUG-2000; 2000US-0225552.
PR 14-AUG-2000; 2000US-0225553.
PR 14-AUG-2000; 2000US-0225554.
PR 14-AUG-2000; 2000US-0225555.
PR 14-AUG-2000; 2000US-0225556.
PR 14-AUG-2000; 2000US-0225557.
PR 14-AUG-2000; 2000US-0225558.
PR 14-AUG-2000; 2000US-0225559.
PR 14-AUG-2000; 2000US-0225560.
PR 14-AUG-2000; 2000US-0225561.
PR 14-AUG-2000; 2000US-0225562.
PR 14-AUG-2000; 2000US-0225563.
PR 14-AUG-2000; 2000US-0225564.
PR 14-AUG-2000; 2000US-0225565.
PR 14-AUG-2000; 2000US-0225566.
PR 14-AUG-2000; 2000US-0225567.
PR 14-AUG-2000; 2000US-0225568.
PR 14-AUG-2000; 2000US-0225569.
PR 14-AUG-2000; 2000US-0225570.
PR 14-AUG-2000; 2000US-0225571.
PR 14-AUG-2000; 2000US-0225572.
PR 14-AUG-2000; 2000US-0225573.
PR 14-AUG-2000; 2000US-0225574.
PR 14-AUG-2000; 2000US-0225575.
PR 14-AUG-2000; 2000US-0225576.
PR 14-AUG-2000; 2000US-0225577.
PR 14-AUG-2000; 2000US-0225578.
PR 14-AUG-2000; 2000US-0225579.
PR 14-AUG-2000; 2000US-0225580.
PR 14-AUG-2000; 2000US-0225581.
PR 14-AUG-2000; 2000US-0225582.
PR 14-AUG-2000; 2000US-0225583.
PR 14-AUG-2000; 2000US-0225584.
PR 14-AUG-2000; 2000US-0225585.
PR 14-AUG-2000; 2000US-0225586.
PR 14-AUG-2000; 2000US-0225587.
PR 14-AUG-2000; 2000US-0225588.
PR 14-AUG-2000; 2000US-0225589.
PR 14-AUG-2000; 2000US-0225590.
PR 14-AUG-2000; 2000US-0225591.
PR 14-AUG-2000; 2000US-0225592.
PR 14-AUG-2000; 2000US-0225593.
PR 14-AUG-2000; 2000US-0225594.
PR 14-AUG-2000; 2000US-0225595.
PR 14-AUG-2000; 2000US-0225596.
PR 14-AUG-2000; 2000US-0225597.
PR 14-AUG-2000; 2000US-0225598.
PR 14-AUG-2000; 2000US-0225599.
PR 14-AUG-2000; 2000US-0225600.
PR 14-AUG-2000; 2000US-0225601.
PR 14-AUG-2000; 2000US-0225602.
PR 14-AUG-2000; 2000US-0225603.
PR 14-AUG-2000; 2000US-0225604.
PR 14-AUG-2000; 2000US-0225605.
PR 14-AUG-2000; 2000US-0225606.
PR 14-AUG-2000; 2000US-0225607.
PR 14-AUG-2000; 2000US-0225608.
PR 14-AUG-2000; 2000US-0225609.
PR 14-AUG-2000; 2000US-0225610.
PR 14-AUG-2000; 2000US-0225611.
PR 14-AUG-2000; 2000US-0225612.
PR 14-AUG-2000; 2000US-0225613.
PR 14-AUG-2000; 2000US-0225614.
PR 14-AUG-2000; 2000US-0225615.
PR 14-AUG-2000; 2000US-0225616.
PR 14-AUG-2000; 2000US-0225617.
PR 14-AUG-2000; 2000US-0225618.
PR 14-AUG-2000; 2000US-0225619.
PR 14-AUG-2000; 2000US-0225620.
PR 14-AUG-2000; 2000US-0225621.
PR 14-AUG-2000; 2000US-0225622.
PR 14-AUG-2000; 2000US-0225623.
PR 14-AUG-2000; 2000US-0225624.
PR 14-AUG-2000; 2000US-0225625.
PR 14-AUG-2000; 2000US-0225626.
PR 14-AUG-2000; 2000US-0225627.
PR 14-AUG-2000; 2000US-0225628.
PR 14-AUG-2000; 2000US-0225629.
PR 14-AUG-2000; 2000US-0225630.
PR 14-AUG-2000; 2000US-0225631.
PR 14-AUG-2000; 2000US-0225632.
PR 14-AUG-2000; 2000US-0225633.
PR 14-AUG-2000; 2000US-0225634.
PR 14-AUG-2000; 2000US-0225635.
PR 14-AUG-2000; 2000US-0225636.
PR 14-AUG-2000; 2000US-0225637.
PR 14-AUG-2000; 2000US-0225638.
PR 14-AUG-2000; 2000US-0225639.
PR 14-AUG-2000; 2000US-0225640.
PR 14-AUG-2000; 2000US-0225641.
PR 14-AUG-2000; 2000US-0225642.
PR 14-AUG-2000; 2000US-0225643.
PR 14-AUG-2000; 2000US-0225644.
PR 14-AUG-2000; 2000US-0225645.
PR 14-AUG-2000; 2000US-0225646.
PR 14-AUG-2000; 2000US-0225647.
PR 14-AUG-2000; 2000US-0225648.
PR 14-AUG-2000; 2000US-0225649.
PR 14-AUG-2000; 2000US-0225650.
PR 14-AUG-2000; 2000US-0225651.
PR 14-AUG-2000; 2000US-0225652.
PR 14-AUG-2000; 2000US-0225653.
PR 14-AUG-2000; 2000US-0225654.
PR 14-AUG-2000; 2000US-0225655.
PR 14-AUG-2000; 2000US-0225656.
PR 14-AUG-2000; 2000US-0225657.
PR 14-AUG-2000; 2000US-0225658.
PR 14-AUG-2000; 2000US-0225659.
PR 14-AUG-2000; 2000US-0225660.
PR 14-AUG-2000; 2000US-0225661.
PR 14-AUG-2000; 2000US-0225662.
PR 14-AUG-2000; 2000US-0225663.
PR 14-AUG-2000; 2000US-0225664.
PR 14-AUG-2000; 2000US-0225665.
PR 14-AUG-2000; 2000US-0225666.
PR 14-AUG-2000; 2000US-0225667.
PR 14-AUG-2000; 2000US-0225668.
PR 14-AUG-2000; 2000US-0225669.
PR 14-AUG-2000; 2000US-0225670.
PR 14-AUG-2000; 2000US-0225671.
PR 14-AUG-2000; 2000US-0225672.
PR 14-AUG-2000; 2000US-0225673.
PR 14-AUG-2000; 2000US-0225674.
PR 14-AUG-2000; 2000US-0225675.
PR 14-AUG-2000; 2000US-0225676.
PR 14-AUG-2000; 2000US-0225677.
PR 14-AUG-2000; 2000US-0225678.
PR 14-AUG-2000; 2000US-0225679.
PR 14-AUG-2000; 2000US-0225680.
PR 14-AUG-2000; 2000US-0225681.
PR 14-AUG-2000; 2000US-0225682.
PR 14-AUG-2000; 2000US-0225683.
PR 14-AUG-2000; 2000US-0225684.
PR 14-AUG-2000; 2000US-0225685.
PR 14-AUG-2000; 2000US-0225686.
PR 14-AUG-2000; 2000US-0225687.
PR 14-AUG-2000; 2000US-0225688.
PR 14-AUG-2000; 2000US-0225689.
PR 14-AUG-2000; 2000US-0225690.
PR 14-AUG-2000; 2000US-0225691.
PR 14-AUG-2000; 2000US-0225692.
PR 14-AUG-2000; 2000US-0225693.
PR 14-AUG-2000; 2000US-0225694.
PR 14-AUG-2000; 2000US-0225695.
PR 14-AUG-2000; 2000US-0225696.
PR 14-AUG-2000; 2000US-0225697.
PR 14-AUG-2000; 2000US-0225698.
PR 14-AUG-2000; 2000US-0225699.
PR 14-AUG-2000; 2000US-0225700.
PR 14-AUG-2000; 2000US-0225701.
PR 14-AUG-2000; 2000US-0225702.
PR 14-AUG-2000; 2000US-0225703.
PR 14-AUG-2000; 2000US-0225704.
PR 14-AUG-2000; 2000US-0225705.
PR 14-AUG-2000; 2000US-0225706.
PR 14-AUG-2000; 2000US-0225707.
PR 14-AUG-2000; 2000US-0225708.
PR 14-AUG-2000; 2000US-0225709.
PR 14-AUG-2000; 2000US-0225710.
PR 14-AUG-2000; 2000US-0225711.
PR 14-AUG-2000; 2000US-0225712.
PR 14-AUG-2000; 2000US-0225713.
PR 14-AUG-2000; 2000US-0225714.
PR 14-AUG-2000; 2000US-0225715.
PR 14-AUG-2000; 2000US-0225716.
PR 14-AUG-2000; 2000US-0225717.
PR 14-AUG-2000; 2000US-0225718.
PR 14-AUG-2000; 2000US-0225719.
PR 14-AUG-2000; 2000US-0225720.
PR 14-AUG-2000; 2000US-0225721.
PR 14-AUG-2000; 2000US-0225722.
PR 14-AUG-2000; 2000US-0225723.
PR 14-AUG-2000; 2000US-0225724.
PR 14-AUG-2000; 2000US-0225725.
PR 14-AUG-2000; 2000US-0225726.
PR 14-AUG-2000; 2000US-0225727.
PR 14-AUG-2000; 2000US-0225728.
PR 14-AUG-2000; 2000US-0225729.
PR 14-AUG-2000; 2000US-0225730.
PR 14-AUG-2000; 2000US-0225731.
PR 14-AUG-2000; 2000US-0225732.
PR 14-AUG-2000; 2000US-0225733.
PR 14-AUG-2000; 2000US-0225734.
PR 14-AUG-2000; 2000US-0225735.
PR 14-AUG-2000; 2000US-0225736.
PR 14-AUG-2000; 2000US-0225737.
PR 14-AUG-2000; 2000US-0225738.
PR 14-AUG-2000; 2000US-0225739.
PR 14-AUG-2000; 2000US-0225740.
PR 14-AUG-2000; 2000US-0225741.
PR 14-AUG-2000; 2000US-0225742.
PR 14-AUG-2000; 2000US-0225743.
PR 14-AUG-2000; 2000US-0225744.
PR 14-AUG-2000; 2000US-0225745.
PR 14-AUG-2000; 2000US-0225746.
PR 14-AUG-2000; 2000US-0225747.
PR 14-AUG-2000; 2000US-0225748.
PR 14-AUG-2000; 2000US-0225749.
PR 14-AUG-2000; 2000US-0225750.
PR 14-AUG-2000; 2000US-0225751.
PR 14-AUG-2000; 2000US-0225752.
PR 14-AUG-2000; 2000US-0225753.
PR 14-AUG-2000; 2000US-0225754.
PR 14-AUG-2000; 2000US-0225755.
PR 14-AUG-2000; 2000US-0225756.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225760.
PR 14-AUG-2000; 2000US-0225761.
PR 14-AUG-2000; 2000US-0225762.
PR 14-AUG-2000; 2000US-0225763.
PR 14-AUG-2000; 2000US-0225764.
PR 14-AUG-2000; 2000US-0225765.
PR 14-AUG-2000; 2000US-0225766.
PR 14-AUG-2000; 2000US-0225767.
PR 14-AUG-2000; 2000US-0225768.
PR 14-AUG-2000; 2000US-0225769.
PR 14-AUG-2000; 2000US-0225770.
PR 14-AUG-2000; 2000US-0225771.
PR 14-AUG-2000; 2000US-0225772.
PR 14-AUG-2000; 2000US-0225773.
PR 14-AUG-2000; 2000US-0225774.
PR 14-AUG-2000; 2000US-0225775.
PR 14-AUG-2000; 2000US-0225776.
PR 14-AUG-2000; 2000US-0225777.
PR 14-AUG-2000; 2000US-0225778.
PR 14-AUG-2000; 2000US-0225779.
PR 14-AUG-2000; 2000US-0225780.
PR 14-AUG-2000; 2000US-0225781.
PR 14-AUG-2000; 2000US-0225782.
PR 14-AUG-2000; 2000US-0225783.
PR 14-AUG-2000; 2000US-0225784.
PR 14-AUG-2000; 2000US-0225785.
PR 14-AUG-2000; 2000US-0225786.
PR 14-AUG-2000; 2000US-0225787.
PR 14-AUG-2000; 2000US-0225788.
PR 14-AUG-2000; 2000US-0225789.
PR 14-AUG-2000; 2000US-0225790.
PR 14-AUG-2000; 2000US-0225791.
PR 14-AUG-2000; 2000US-0225792.
PR 14-AUG-2000; 2000US-0225793.
PR 14-AUG-2000; 2000US-0225794.
PR 14-AUG-2000; 2000US-0225795.
PR 14-AUG-2000; 2000US-0225796.
PR 14-AUG-2000; 2000US-0225797.
PR 14-AUG-2000; 2000US-0225798.
PR 14-AUG-2000; 2000US-0225799.
PR 14-AUG-2000; 2000US-0225800.
PR 14-AUG-2000; 2000US-0225801.
PR 14-AUG-2000; 2000US-0225802.
PR 14-AUG-2000; 2000US-0225803.
PR 14-AUG-2000; 2000US-0225804.
PR 14-AUG-2000; 2000US-0225805.
PR 14-AUG-2000; 2000US-0225806.
PR 14-AUG-2000; 2000US-0225807.
PR 14-AUG-2000; 2000US-0225808.
PR 14-AUG-2000; 2000US-0225809.
PR 14-AUG-2000; 2000US-0225810.
PR 14-AUG-2000; 2000US-0225811.
PR 14-AUG-2000; 2000US-0225812.
PR 14-AUG-2000; 2000US-0225813.
PR 14-AUG-2000; 2000US-0225814.
PR 14-AUG-2000; 2000US-0225815.
PR 14-AUG-2000; 2000US-0225816.
PR 14-AUG-2000; 2000US-0225817.
PR 14-AUG-2000; 2000US-0225818.
PR 14-AUG-2000; 2000US-0225819.
PR 14-AUG-2000; 2000US-0225820.
PR 14-AUG-2000; 2000US-0225821.
PR 14-AUG-2000; 2000US-0225822.
PR 14-AUG-2000; 2000US-0225823.
PR 14-AUG-2000; 2000US-0225824.
PR 14-AUG-2000; 2000US-0225825.
PR 14-AUG-2000; 2000US-0225826.
PR 14-AUG-2000; 2000US-0225827.
PR 14-AUG-2000; 2000US-0225828.
PR 14-AUG-2000; 2000US-0225829.
PR 14-AUG-2000; 2000US-0225830.
PR 14-AUG-2000; 2000US-0225831.
PR 14-AUG-2000; 2000US-0225832.
PR 14-AUG-2000; 2000US-0225833.
PR 14-AUG-2000; 2000US-0225834.
PR 14-AUG-2000; 2000US-0225835.
PR 14-AUG-2000; 2000US-0225836.
PR 14-AUG-2000; 2000US-0225837.
PR 14-AUG-2000; 2000US-02

PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249254.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251719.
PR	06-DEC-2000;	2000US-0251719.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-488783/53.	
DR	N-PSDB: AAS26613.	
XX		
PT	New nucleic acid molecules encoding 461 human secreted proteins for	
PT	diagnosing, preventing, treating or ameliorating medical conditions and	
PT	used as food additives or preservatives -	
XX		
PS	Claim 11; SEQ ID NO 1579; 980pp; English.	
XX		
CC	The invention relates to isolated nucleic acid molecules and their	
CC	encoded secreted proteins. The nucleic acids and proteins are used to	
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,	
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They	
CC	are also used in diagnosing a pathological condition or susceptibility	
CC	to a pathological condition. Antibodies to the proteins can also	
CC	be used in alleviating symptoms associated with the disorders and in	
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	
CC	immunosorbant assays (ELISAs). Disorders which are diagnosed or treated	
CC	include autoimmune diseases e.g. rheumatoid arthritis,	
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,	
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	

CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match	96.7%;	Score 2972.5;	DB 22;	Length 718;
Best Local Similarity	99.1%;	Pred. No. 2.2e-239;		
Matches 573;	Conservative	0;	Mismatches 4;	Indels 1;
				Gaps 1;

QY	3	KKKKKKSDKHLVEEYVYKPLKTLVKYGVGNVPTLSGSSGSDHSIPEEDKNDNDKHKDK	62
Db	70	KKKKKKSDKHLVEEYVYKPLKTLVKYGVGNVPTLSGSSGSDHSIPEEDKNDNDKHKDK	129
QY	63	KKKKKKGEKQIQIEEGEKKRRRRVYEDKKKRDREVENEAEDKQLQCAVPRLDPEPKPLT	122
Db	130	KKKKKKGEKQIQIEEGEKKRRRRVYEDKKKRDREVENEAEDKQLQCAVPRLDPEPKPLT	189
QY	123	SSLAQEEVEQPTQLQALNQLRDLQKRDPAFSPVTPDIAFGYMIKKHMDPSTMK	182
Db	190	SSLAQEEVEQPTQLQALNQLRDLQKRDPAFSPVTPDIAFGYMIKKHMDPSTMK	249
QY	183	EKIKNNNDYQSIEELKDNKLMCTNAMIYNKPEITYYRAAKKLHSGMKILSQERLQSLQ	242
Db	250	EKIKNNNDYQSIEELKDNKLMCTNAMIYNKPEITYYRAAKKLHSGMKILSQERLQSLQ	309
QY	243	STDFEADLQKRRKQKGDGNTDSQSGDEGCGWREREDSGDAFAHAFKSPKKNKKDKML	302
Db	310	STDFEADLQKRRKQKGDGNTDSQSGDEGCGWREREDSGDAFAHAFKSPKKNKKDKML	369
QY	303	EDKFKSNMLREBQELDRIVKESGGKTLRRLVNSQCEFERRRKPDGTTGLLHPDPIVG	362
Db	370	EDKFKSNMLREBQELDRIVKESGGKTLRRLVNSQCEFERRRKPDGTTGLLHPDPIVG	429
QY	363	EEGCVCLVRIKMTGTGLOGGVNLTQEFKKDKNKTKTPVLYLWGVGYSSAPHYDSTFANIS	422
Db	430	EEGCVCLVRIKMTGTGLOGGVNLTQEFKKDKNKTKTPVLYLWGVGYSSAPHYDSTFANIS	489
QY	423	KDDSDLIYSTVGEDDLDPDSFIHFFLATCCODPYVMADSLDLVLTGKHSRTLQENEMS	482
Db	490	KDDSDLIYSTVGEDDLDPDSFIHFFLATCCODPYVMADSLDLVLTGKHSRTLQENEMS	549
QY	483	LPEDSGHTTDTGCKEMQITVEVPPGRLDSSDQRLALAAVYNFGVPPVPPSEAEI	542
Db	550	LPEDSGHTTDTGCKEMQITVEVPPGRLDSSDQRLALAAVYNFGVPPVPPSEAEI	608
QY	543	FOKKLDETRRLRELOEAQNERLSTRRPGNMICLLGPS	580
Db	609	FOKKLDETRRLRELOEAQNERLSTRRPGNMICLLGPS	646
RESULT 4			
AAM93712			
ID	AAM93712 standard; Protein; 405 AA.		
AC	AAM93712;		
NC			
XX	06-NOV-2001. (first entry)		
XX	Human polypeptide, SEQ ID NO: 3652.		
XX	Human, full length cDNA, cDNA synthesis; oligo-capping.		
OS	Homo sapiens.		
PN	EPI130094 -A2.		

```

XX 05-SEP-2001.
PD
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
XX N-PSDB: AAK94661.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 3652; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesized by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 405 AA:
SQ

```

Query Match 55.7%; Score 1710.5; DB 22; Length 405;
 Best Local Similarity 98.8%; Pred. No. 2,6e-134;
 Matches 330; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

OY 247 MADLQKRRKRCQCTDMSQSGEDGCGWOREDSGDAEAHAFKSPSKENKKRDKMDLEDKF 306
DB 1 madqkrrkrcqctdmsqsgedgcgworedsgdaeahafkspskennkkdkmdledkf 60
OY 307 KSNMLERQDQLDRIVKESGKLTRLVNSQCEFERRRKPDGTTGLHHPVDIVCEPGT 366
DB 61 ksnmlereqqldrivkesgklttrlvnsqceferrrkpdgttllgllhpvdivepgy 120
OY 367 CLVRLGNTGRLOSGVNTLOGFEKEDKRNKVTPLYLNTGYSYAPHYDSTFANISKDS 426
DB 121 clvrlgntgrlqsgvntlogfekedkrnkvtplylntgysyaphydstfanskds 180
OY 427 DLVSYRYGSDSLPSPFSIHFLATCQDYPYVADSLDLVLRKGGHSRTLQEMENSLPND 486
DB 181 dlvsyrygsdslpsfshflatcqdyvmadslldlvlrkgsrslqememslpnd 240
OY 487 EGHTRRLDTGKEMEQITTEVEPPGRLLDSTQDRLIALKAVTNGVPEVDSERAEFFQK 546
DB 241 eghtrrltdtgkemeqitteveppgrlldstqdrllalkavtngvpevdseaeelfqk 299
OY 547 LDETTIRLRLKLOEAOERLSTRPGNMICLLGPS 580
DB 300 ldettrrlrlkloeaoerlstrpgnmicllgps 333

```

RESULT 5
 AAB93765 standard; Protein: 351 AA.
 XX AAB93765;
 AC
 XX

```

DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13461.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13461; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 351 AA:
SQ

```

Query Match 46.0%; Score 1413.5; DB 22; Length 351;
 Best Local Similarity 98.2%; Pred. No. 1.3e-109;
 Matches 275; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```

OY 301 MLEDFKSNMLERQDQLDRIVKESGKLTRLVNSQCEFERRRKPDGTTGLHHPDPI 360
DB 1 mledfkksnmlereqqldrivkesgklttrlvnsqceferrrkpdgttllgllhpvdp 60
OY 361 VGEFGYCLVRLGNTGRLOSGVNTLOGFEKEDKRNKVTPLYLNTGYSYAPHYDSTFAN 420
DB 61 vgefygclvrlgntgrlqsgvntlogfekedkrnkvtplylntgysyaphydstfan 120
OY 421 ISKDSDLIVSYGSDSLPSPFSIHFLATCQDYPYVADSLDLVLRKGGHSRTLQEME 480

```

Db 121 iskdssdlystygadedslpsdisihflatcdqdyymadsildivltkgshstlqeme 180
 QY 481 MSIPEDSGHFTLTPTGKMEQITVEPPGRLDSTQDRLTAKAVTNFVGVEVDESEA 540
 Db 181 mslypdeghrtltakeme-iteveppgrldstqdrllalkavtnfygvvefdeaea 239
 QY 541 EIFQKILDETRRLRELQEAONERLSTPPGNMTCLLGPS 580
 Db 240 eifgkildetrllrelqeaqnerlstrppgmictllgps 279
 RESULT 6
 AAM38835
 ID AAM38835 standard; Protein: 597 AA.
 AC AAM38835;
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 1980.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.
 OS Homo sapiens.
 PN MO200153312-A1.
 PD 26-JUL-2001.
 PF 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI57991.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 3; SEQ ID NO 1980; 10078pp; English.
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SQ Sequence 597 AA;
 Query Match 30.0%; Score 921.5; DB 22; Length 597;
 Best Local Similarity 34.7%; Pred. No. 3,8e-68;
 Matches 213; Conservative 127; Mismatches 184; Indels 89; Gaps 17;
 QY 1 MGKKKKKKSD-KHLYEEY---VEKPLKLVLYGVGNEVELSTGSSGHDSTFDEKKNH 55
 Db 1 mgkhhkhhkaewrsyedyadkplkplklylvvgsevvels--gshsssyddrshd 58
 QY 56 --DKHKRRKKRRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKRKKR 110
 Db 59 erethkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrrkkrr 115
 QY 111 VRDLPEPKPLTSSIAKQEEVEQTPLOEALNQLMRQIQRKPSAFSPVTFPIAGYSM 170
 Db 116 vevppdrpractqpaenestpiqllhflrqlqrkdpngfifaivtdalapygm 175
 QY 171 IIRKPNQFSTYKKEIKRNDQSTIELKDNFKIMQTNMIVNKPEITYYKAKKLLSGMK 230
 Db 176 ilkhpmdfgltmkdklvaneysvlefakdiklmednamynrptvykylakllhagfk 235
 QY 231 ILISERI-----OSLKQSIDFMADLQTKRKQKDGTDTSQSGEDGCGWQREDSGDAA 284
 Db 236 mmskqaaallgnedaveepvepvq----- 262
 QY 285 HAFKSPSKENKKKDKMDLEDFK-----SNNLERQEQDLRIVKGSGKLTFRRLV 334
 Db 263 --vetakskkpsrevyscmfdepgnacsldstakehvalavehaeardrinfllp 319
 QY 335 NSOCEFERKRPDGTTLGLHPVPIVGEPPGYCLVRLGMTTGRLOSQSVNTLQGFEDKRN 394
 Db 320 gqkngylkrngdgellysvntlaepdadeethpvdslsskllppfttl-gfkderm 378
 QY 395 KVTPLYVINYGPYSSYAPHYDSTFANISKDDSLIYSTYGDSLPBDFSHFELACQD 454
 Db 379 kvt--fls-sattalsmqnsvfgdtksdemellyasgdeqlvgalslqetlvkags 434
 QY 455 YPYVADSLDLVLRKGSSTFLOEME-----MSLDEGHTRL-PTGKMEQITVEP 507
 Db 435 yskrvvddlllqitgdstlrlfqlkqrrnypmnpdeakvqdtlqgsssvlefmsmk- 493
 QY 508 PGRLDSTQDRLTAKAVTNFVGVEVDESEAETFOKKILDETRRLRELQEAONERLST 567
 Db 494 -----sydvsvdismsslgkvkkelppdshl---nldetkllqldlneagaerygs 544
 QY 568 RPPGNMTCLLGPS 580
 Db 545 rpsnlslslnas 557
 RESULT 7
 AAM40621
 ID AAM40621 standard; Protein: 605 AA.
 AC AAM40621;
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 5552.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0233081.
 PR 12-SEP-2000; 2000US-02331968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0234400.
 PR 14-SEP-2000; 2000US-0234401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249287.
 PR 17-NOV-2000; 2000US-0249289.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-488783/53.
 DR N-PSDB; AAS26200.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PT
 XX
 PS Claim 11; SEQ ID NO 1166; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention.

Query Match 26.0%; Score 798.5; DB 22; Length 199;
 Best local Similarity 41.1%; Pred. No. 1.4e-58;

Matches 176; Conservative 0; Mismatches 1; Indels 251; Gaps 1;

```

QY 17 EYVEKPLKLVKVGMEVTELTSGSSGHSLSFEDKNDHKDKRRKKRKGEOIPGE 76
    |||||
Db 1 eyekpklklykvgnevtelstgssghsdslfedkndhdkhdkrrkkkgekpge 60
QY 77 EKGRKRRVAKKKKRDROVNEAEKDLQCHAPVRLDLPPEKPLTSLAKOEVEQTP 136
    |||||
Db 61 ekgrkrrv-----
QY 137 OEALNQLMRQLQKRPSPFSPVTDIARGYSMITKHPMDFSTMKKIKNNYQSTIEL 196
    |||||
Db 71 -----
QY 197 KDNFKLMCTNAMIYKRPETIYKAAKLLHSGMKILSOERIOSKSIDFMADLOKTRKQ 256
    |||||
Db 71 -----
QY 257 KDGTDTSGSGEDGCGNOREDESDAEAHAFKSPSKENKKKDKMDLEDKFSNNLEBDE 316
    |||||
Db 71 -----
QY 317 QLDRIVESGKLTLRRLVNSOCEFEERRKPDGTTTGLHPVDPIVGPYCLVRLGNTTG 376
    |||||
Db 71 -----ceferrrkpdgttllglhpvdpivgpycprlqmttg 109
QY 377 RLQSGVNTLOGFEKEDKRNKVTPLYLYNGPYSSYAPHYSTFANISKDSLLYSTYGED 436
    |||||
Db 110 rllsgvntllqgfekdrnkvtplyllyngpyssyaphystfanskdsdlystyged 169
QY 437 SGLPSDPS 444
    |||||
Db 170 sdllpsdps 177

```

RESULT 9

ABB63028 standard; Protein; 861 AA.

AC ABB63028;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 15876.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07131.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 XX Disclosure; SEQ ID NO 15876; 21np + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 861 AA;

Query Match 22.8%; Score 699.5; DB 22; Length 861;
 Best local Similarity 27.6%; Pred. No. 2.3e-49;

Matches 211; Conservative 127; Mismatches 219; Indels 207; Gaps 22;

```

QY 3 KKKKKKSD-KHLYEY-----VEKPLKLVKVGMEVTELTSGSSGHD----- 45
    |||||
Db 5 kkkkkkserleeyeeysghqpaqlrgllklllygsnatpaysanpmvqpprlae 64
QY 46 ---SSLFEDKNDHKDKRRKKRKGEOIGEEKRRARRVKEKKRDROVNEAE 102
    |||||
Db 65 ammspvpeelqdgqhnrhkkkkskkk---kdrkhhkhke-krnsrdhrdags 119
QY 103 -----KDLOCH----- 108
    |||||
Db 120 dedmaggadaacsqfapsvappaadpsqsgdgsfmdddsgqrlpenllffaglttch 179
QY 109 -----APRLD-----LP----- 116
    |||||
Db 180 spsnpcvtklpaprklldllmgsspsusslqsslqllgspckrlpdlpssptp9ga 239
QY 117 -----PEKPLTSLAKOE-----EVEQTPLOEALNQLMRQLQKRPSPF 155
    |||||
Db 240 nslnahtpkaleapkrpssssssgrepscvllklyqkspnlkllhllrflckrpbqf 299
QY 156 FSPVYDFTAPGYSMTIKHPMDFSTMKKIKNNNDYOSIEELDNFKLMCTNAMIYKRPET 215
    |||||
Db 300 fawpyrddmappysllsrpmdfstmrqklddheytaletfdffklmcenaikynhndt 359
QY 216 IYKKAARKLLHSGMKILSOERI-QSLKOSIDPMADLOKTRKOKDGTDTSGSGEDGCGMOR 274
    |||||
Db 360 vynkaakrlllygmkhlpgeenlmrelkplsgymrel---tarelfeljsnd-----tar 411
QY 275 EREDSGD-----AEAHAKSPSKENKKKDKMDLEDKFS-----NNLEERG--EQLDR 320
    |||||
Db 412 emndsdegastgaetpraqleeertrclrltenaprlthfepvydltlgeellaqyn 471
QY 321 IYKESGKLTLRRLVNSOCEFEERRKPDGTTTGLHPVDPIVGPYCLVRLGNTTGRLQS 380
    |||||
Db 472 aagqakgrvnaaknhkngflrkmkgdgttllnlvkee---negpeyvtldllygklqk 528
QY 381 GYNTILOGFEKEDKRNKVTPLYLYNGPYSSYAPHYSTFANISKDSLLYSTYGEDSLP 440
    |||||
Db 529 gsaqltqrqvdkrnavrvtkslnygaafasfapclfsrlfaetqylvrtygassaae 588
QY 441 SDFSIEHFLATCODYPYVWADSLDLVLTGCGHSRTL-----QEHENSLPDE 487
    |||||

```

Db 589 yaesllqftkd-snygtiangllidllngeshskldelynmqlhyeqrelekcfegee 647
 Oy 488 GHTRTLDTGKMEQITTEVEPPGRLDSSSTODRLALKAATNCGVPEVDESEAEI----- 542
 Db 648 -----eessggettaqiegeyekyknthvdfkqisigeligidvstfidmeaemksyel 701
 Oy 543 ---FQKRLDETRRLRELQEAQNERLSTRPGNMICLLGSPSEK 583
 Db 702 nrmhehlsqnltlleklirvaghdrls-qplpnhlglvqpagge 744

RESULT 10
 AAB95421
 ID AAB95421 standard; Protein: 501 AA.
 AC AAB95421;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17823.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PE 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 17823; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 501 AA;
 Query Match 22.3%; Score 684; DB 22; Length 501;
 Best Local Similarity 34.7%; Pred. No. 2e-48;
 Matches 164; Conservative 93; Mismatches 164; Indels 52; Gaps 11;

Oy 130 EVEQTPILQEAALNQLROLQKRDPSAFSPFPTDILAGYSMTIKHPMDFTMKKIRND 189
 Db 19 enestpiqqllehlrlqqrkdpgffafpytdalapgysmlkhpndftmkdkivane 78
 Oy 190 YQSI BELKDNFKMCTNMAIYKRPETIYYKAAKKLHSGMKILSOERIQSISIDPMAD 249
 Db 79 yksvtefkadfklnchdamctyupdtvykklakllhagfkmskerllaiktsmsfmqd 138
 Oy 250 LQKTRKOKDPTDTSGSGEDGCMQREPRDSGDAPAHAFKSP-----SKENKKKDKDLEED 304
 Db 139 m-----disgqaallg-----nedlaveepvpevpvgyelakskkperevlsc 183
 Oy 305 KFK-----SNLTEREQEQLDRIVKESGKILTRLVNSQCFRRKRPDGTTLGLL 354
 Db 184 mfepegnacsltdstaehvalavehaadeardrinfllpgkmgylkrngdgsllysv 243
 Oy 355 HPVDPYGEPEYCLVRLGAMTGRLOSGVNTLOGKEDKRNKVTPLYLNTGPPSYAPHY 414
 Db 244 ntaepdadeeethpvdlsisskllpgfttl-gfkdeerrnkvtf--fls--sattalmsqn 298
 Oy 415 DSTFANISKDSDSLIYSGYGEDSLPSPDSIHEFLANCOQDPYMAADSLDVLKGGHSR 474
 Db 299 nsvlgdlsdemeallysaagdeivgcalslqetvkdaagsykvvdlllqqlvgdgsnr 358
 Oy 475 TLOEME-----MSLPDEDEGHTRTL-DTGKDEQITTEVEPPGRLDSSSTODRLALKAATN 527
 Db 359 tlfdlqgrnympkpdcakgvdtclgdsessvlefmusk-----sydpvsvdlmss 411
 Oy 528 FGVEVEVDESEAEI FQKRLDETRRLRELQEAQNERLSTRPGNMICLLGSPS 580
 Db 412 lqkvkkelldpddshl---nidetklllqdlheaqaergysrpsnlslnas 461

RESULT 11
 AAB41780
 ID AAB41780 standard; Protein: 280 AA.
 AC AAB41780;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1544 polypeptide sequence SEQ ID NO:3088.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vunerary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteoprotic; antiarthritis; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergic; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.


```

XX  Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI  Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI  Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX  WPI; 2001-442253/47.
DR  N-PSDB; AAI60173.
XX
PT  Novel nucleic acids and polypeptides, useful for treating disorders
PT  such as central nervous system injuries -
XX
PS  Example 2; SEQ ID NO 5948; 10078bp; English.
XX
XX  The invention relates to human nucleic acids (AA157798-AA161369) and
CC  the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
CC  immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC  in gene therapy. A composition containing a polypeptide or polynucleotide
CC  of the invention may be used to treat diseases of the peripheral nervous
CC  system, such as peripheral nervous injuries, peripheral neuropathy and
CC  localised neuropathies and central nervous system diseases, such as
CC  Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC  lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC  utilisation of the activities such as: Immune system suppression,
CC  Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC  and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC  assays for receptor activity, arthritis and inflammation, leukaemias and
CC  C.N.S disorders.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification.
XX
SQ  Sequence 1109 AA;

```

```

XX  AAM41018;
AC  22-Oct-2001 (first entry)
DT  Human polypeptide SEQ ID NO 5949.
DE
XX
XX  Human; neurotrophic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX  peripheral nervous system; neuropathy; central nervous system; CNS;
XX  Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX  amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX  chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX  leukaemia.
XX
XX  Homo sapiens.
OS
XX  WO200153312-A1.
XX
XX  26-JUL-2001.
PD
XX
XX  26-DEC-2000; 2000MO-US34263.
PF
XX
XX  21-JAN-2000; 2000US-0488725.
PR
XX  25-APR-2000; 2000US-0552317.
PR
XX  09-JUL-2000; 2000US-0598042.
PR
XX  19-JUL-2000; 2000US-0620312.
PR
XX  03-AUG-2000; 2000US-0653450.
PR
XX  14-SEP-2000; 2000US-0662191.
PR
XX  19-OCT-2000; 2000US-0693036.
PR
XX  29-NOV-2000; 2000US-0727344.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI  Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI  Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX  WPI; 2001-442253/47.
DR  N-PSDB; AAI60174.
XX
XX  Novel nucleic acids and polypeptides, useful for treating disorders
XX  such as central nervous system injuries -
XX
PS  Example 2; SEQ ID NO 5949; 10078bp; English.
XX
XX  The invention relates to human nucleic acids (AA157798-AA161369) and
CC  the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
CC  immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC  in gene therapy. A composition containing a polypeptide or polynucleotide
CC  of the invention may be used to treat diseases of the peripheral nervous
CC  system, such as peripheral nervous injuries, peripheral neuropathy and
CC  localised neuropathies and central nervous system diseases, such as
CC  Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC  lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC  utilisation of the activities such as: Immune system suppression,
CC  Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC  and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC  assays for receptor activity, arthritis and inflammation, leukaemias and
CC  C.N.S disorders.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification.
XX
SQ  Sequence 1109 AA;

```

```

Query Match      8.4%; Score 258.5; DB 22; Length 1109;
Best Local Similarity 23.5%; Pred. No. 2,3e-12;
Matches 109; Conservative 62; Mismatches 136; Indels 157; Gaps 15;

```

```

Query Match      8.4%; Score 258.5; DB 22; Length 1109;
Best Local Similarity 23.5%; Pred. No. 2,3e-12;
Matches 109; Conservative 62; Mismatches 136; Indels 157; Gaps 15;

```

```

OY 71 K-----QIPGEEGRKRRRVKEDKK-----91
Db 486 kalaepcavlpvcapyipqrlnrldanqvaiaqrkkqfverahsywlkrlsrngapllr 545
OY 92 -----RDRDRVENEAEKDLOCHAPVRLDLPPEKPYTSSIAKOE-----E 130
Db 546 rlgslqsgqrssqgrendeemkaakekikwyqrlrhdlerarlllellrkreklkregvk 605
OY 131 VEO-----TPLOEALNQLMRQLORKDPsAFSEpVYDFIApGYSMTIKHPMDFSTIKE 183
Db 606 veqvamelrlplvlrsvldqldkqparlfagpvsikevpdyldhikhpmdfatmrk 665
OY 184 KIKNDYOSIEELKDNFKIMCTNAMITYNKPEITYYKAAKRLHSGMKILSQERIOSLKOS 243
Db 666 rleaggyxnlhefeedfaldldncmkynardtvlryaavrlrdggyvllrqar-----718
OY 244 IDFMADIAKTRKQKDGTPDSQSGEDGCGWOREREDSGDAAE--HAKSPSKENKK----296
Db 719 -----revdsigleasgmhlperpaaprrpfsw 748
OY 297 KDKDMLEDKFKSNL---EREOJDRI-----VKESGKILR--RLVNSQCEFERRKPDG 347
Db 749 edvdrllpnanrahlgleeqlreilldldlcamksgrsrkrakllkkaiaallrnk---805
OY 348 TTTTGLHPVDPIVGPBGYCLVRLGMTGRLQSGVNTLQGFRED 391
Db 806 ---leqghs-qplptgpg-----legfeed 826

```

Search completed: July 11, 2002, 15:49:34
 Job time: 1023 sec

Align seg 1/1 to: BRF3_HUMAN from: 1 to: 1214

```
74 CGGCGCTCGGCTGGCCCGGACCGGAGCGGCGGCGCGCTGG 123
|||||
397 ProPoglyAlaIaIaThrAlaArgArgLysGlyAspSerProAlaSerI 413
|||||
124 GCGTCGCGCGGGGGCGGACCGGGGCGGCTCGACATGGG..... 166
|||||
413 eSerGIuThrGIuAspGIuGIuGIuLysGIuGIuGIuGIuGIuGIu 430
|||||
167 .....CAAGAACACAAAGAACACAAAGTCGGACAA 196
|||||
430 LuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 446
|||||
197 ACA...CCCTACGAGAGTA..... 214
|||||
447 SerGIuSerLeuLysGIuAlaProLysLysSerLysMetSerLeuLysGI 463
|||||
215 .....TGTAAGAAAGCCCTGAAGCTGGT..... 238
|||||
463 nLysIleLysLysGIuProGIuGIuAlaGIuLysnAspThrProSerThrL 480
|||||
239 .....CCTCAAGTAGAGGAGAACGAATCACCGAGCT 271
|||||
480 euProMetLeuAlaValProGIuIleProSerTyrArgLeuAsnLysIL 496
|||||
272 CTCACAGGCGACCTCGGGCGACGACTCCAGCTCTTCGAAGACAAAGAG 321
|||||
496 eCysSerGIuLeuSer.....PheGIuArgLysAsnG 507
|||||
322 ATCATGACAAACACAAAGACAGAAAGCGGAAAGAAAGAAAGAGAGAG 371
|||||
507 LnPheMetGIuArgLeuHisAsnTyrTrpLeuLysArgLysnAlaArg 523
|||||
372 AAGCAGATTCCAGGGGAGAAAGGAGAAAGAGAGAGAGTAAGGA 421
|||||
524 AsnGIuValProLeuIleArgArgLeuHisSerHisLeuGIuSerGIu 540
|||||
422 GGATAAAGAAAGACGAGATCGAGACCGGGTGAGAAATGAGCAGAAAG 471
|||||
540 GAsnAlaGIuGIuArgGIuGIuAspGIuThrSerAlaValLysGIuG 557
|||||
472 ATCTCAGTGTCACGCGCTGTGAGATTAGACTTGCCTCTGAGAAAGCT 521
|||||
557 LuLeuLysTyrTrpGIuLysLeuArgHisAspLeuGIuArgAlaArgLeu 573
|||||
522 CTCACAGAGCTTTAGCCAAACAGAGAGAGTA.....GAACAGACAGC 565
|||||
574 LeuIleGIuLeuIleArgLysArgGIuLysLeuLysArgGIuGIuVal 590
|||||
566 CCTTCAGAGAGCT.....TTGAATCAACTGA 591
|||||
590 sValGIuGIuAlaIaMetGIuLeuGIuLeuMetProPheAsnValLeu 607
|||||
592 TGAGACAA.....TTGCAGAGAAAGATCCAAAGTCTTCTT 629
|||||
607 euArgThrThrLeuAspLeuGIuGIuLysAspProAlaHisIlePhe 623
|||||
630 TCATTTCCCTGACATGATTTATGTCTCTGCTACTCCATGATCACTTAA 679
|||||
624 AlaGIuProValAsnLeuSerGIuAlaProAspTyrLeuGIuPheIle 640
|||||
680 ACACCAATGATTTTATGATCATGAAAGAAAGATCAAGAAATGATGCT 729
|||||
640 rLysProMetAspPheSerThrMetArgLysLeuGIuSerHisLeu 657
|||||
730 ATCAGTCCATAGAGACTAAAGATACTCAAACTAATGTGTACTAAT 779
|||||
657 yArgThrLeuGIuGIuPheGIuGIuAspPheAsnLeuIleValThrAsn 673
|||||
780 GGCATGATTTTAAATTAACAGACAGACATTTTATTAAGCTGCAAGAG 829
|||||
674 CysMetLysTyrAsnAlaLysAspThrIlePheHisArgAlaIaVal 690
```

```
830 GCTGTTCACATCAGAGATGAAATTTCTTACCCAGGAAAGATTCAGAGCC 879
|||||
690 GLeuArgAspLeuGIuGIuAlaIaIleLeuArgHisAlaArgArgGIuAla 707
|||||
880 TGAGACAGAGCATAGACTTCATGGCTGACTTGCAGAAAACCTGAAAGCAG 929
|||||
707 LuAsnIleGIuTyrAsp.....Pro 713
|||||
930 AAAGATGAGACAGACACCTCACAGAGTGGGAGACGAGGCTGTGAGCA 979
|||||
714 GIuArgGIuThrHisLeuProGIuSerProLysLeuGIuAspPheTyrAr 730
|||||
980 GAGAGAGAGAGAGACTCTGGAGAT.....GCCAGACACAGCGCT 1020
|||||
730 gPheSerTrpGIuAspValAspAsnIleLeuIleProGIuAsnArgAla 747
|||||
1021 TCAGAGTCCCGACAGAAAGAAATTAAGAAAGACAAAGATATGCTTGA 1070
|||||
747 IsLeuSerProGIuValGIu..... 753
|||||
1071 GATAGTTTAAAGCAATATTTAGAGAGAGAGCAGAGACGCTTGACCG 1120
|||||
754 .....LeuLysGIuLeuGIuLysLeuAspPLe 763
|||||
1121 CATC.....GTGAAGAAATCTGAGAGAAAGCTGACAGG...CGGCTTG 1161
|||||
763 uValSerAlaMetArgSerSerGIuAlaArgThrArgArgValArgLeuL 780
|||||
1162 TGACACATCAGTGCAGATTTGAAAGAAAGAAACCGATGGAACACAGCAG 1211
|||||
780 euArgArgGIuLysnAlaLeuArgLys..... 790
|||||
1212 TTGGGACTTCTCCATCTGTGGATCCATGTAGAGAGACCGAGCTACTG 1261
|||||
791 .....LeuAlaGIuProProProPro..... 797
|||||
1262 CCGTGAGACTGGGAGATGACAACTGGAAGACTTCAGTGTGAGTGAATA 1311
|||||
797 ..... 797
|||||
1312 CTTTCAGAGGGGTTCAAGAGATTAAGAAACAAAGTCACTCCAGTGTTA 1361
|||||
798 .....GlnProProSerLeuAsnLysThrValSerAsnGIu 809
|||||
1362 TATTGAATTTATGGCCCTTACAGTCTTATGCACCGCATTTATGATCCAC 1411
|||||
810 GIuLeuProAlaGIuProGIuGIuLysAsp...AlaAlaValLeuGIuGIuAl 825
|||||
1412 ATTTGCAATATACAGCAAGAGATGATTCGATTTAATCTATTCACCTATG 1461
|||||
825 AlLeuGIuGIuGIuProGIuAspAspGIuLysAsp.....A 836
|||||
1462 GGGAGAGCTGTGATCTCCAAAGTGAATTCAGACATCATGATTTTGGCC 1511
|||||
836 rGAspAspSerLysLeuProProPro..... 844
|||||
1512 ACGTCCCAAGATTATCCGTATGTCATGCGACAGATAGTTACTGATGTTT 1561
|||||
845 .....ProThrIle 847
|||||
1562 AACAAAGAGGGCATTCACAGACCTTACAGAGATGAGATGTATTCG 1611
|||||
847 uGIuProThrGIuProAlaProSerLeuSerGIuGIuGIu...SerProp 863
|||||
1612 CTGAGATGAGAGCCATACAGACATTCACACAGAGAAAGAAATGAGG 1661
|||||
863 rGIuLPro.....ProThrLeuLys 869
|||||
1662 CAGATTACAGAGATGAGACACAGCGGCTTTGACTCCAGTACTCAAGA 1711
|||||
870 ProIleAsnAspSerLysProProSerArg..... 879
```



```

588  AAlaGlnProValSerLeuIysGluValProAspTyrLeuAspHisIleIys 604
680  AACACCAAGATTTTATAGCATGAAGAAAGATCAAGAACATGACT 729
604  SHSPrometAspPheAlaThrMetArgLysArgLeuGluAlaGlnGlyT 621
730  ATCACTCCATAGAACACTAAAGATTAAGTCAAACTAATGTCTACTAAT 779
621  yrltysmleuHnsgluPheGluGluAspPheSpleulleleAspAn 637
780  GCCATGATTACATTAACACGAGACACTTTATTAAGCTGCAAGAA 829
638  CysMetLysTyrAsnAlaArgAspThrValPheTyrArgAlaAlaValAr 654
830  GCGTTCGACATCAGAAATGAAATTTCTAGCCAGAAAGAAATTCAGAGC 879
654  glenutArgAspGlnGlyGlyValValLeuArgGlnAlaArg 667
880  TGAAGCAGACATAGACTTCATGGCTGACTTCAGAAATCGAAACAG 929
667  ..... 667
930  AAAGATGACAGACACTCAGAGAGTGGGAGAGAGAGCTGCTGGCA 979
667  ..... 667
980  GAGAGAGAGAGAGACTCGAGATCGCGAAGA.....CAGCGCT 1020
668  ArgGluValAspSerIleGlyLeuGluGluAlaSerGlyMetHisLeuP 684
1021  TCAAGAGTCCCGCAAGAAATATAAAG.....AAGACAAA 1058
684  roGluArgProAlaAlaAlaIaIaProArgArgProPheSerTrpGluAspVal 700
1059  GATATCTGTGAGATTAAGTTAAACCATTAATTA.....GAGAG 1099
701  ASPATGLeuLeuAspProAlaAsnArgAlaHisLeuGlyLeuGluGlu 717
1100  AGAGCAGAGCAGACTTGACCGCATC.....GTGAAGAA 1134
717  nleuArgGluLeuLeuAspMetLeuAspLeuThrCysAlaMetLysSerS 734
1135  CTGGAGAGAAAGCTGACACAG...CGGCTTGACAGCTAGCGCAATTT 1181
734  erGlySerArgSerLysArgAlaLysLeuLeuLysGluLeuAlaLeu 750
1182  GAAAGAGAAAGAACCATGATGGAACAGAGCTTGCGACTTCTCATCTGT 1231
751  LeuArgAsnLys.....LeuSerGlnGlnHisSer.. 760
1232  GGATGCCATTTAGAGAGACCGAGCTACTGCGGTGAGAGCTGGGAATGA 1281
761  GlnProLeuProThrGlyProGly..... 768
1282  CAACCTGGAAGACTTCAGTCTGGAGTCAATACTTTCAGAGGGTCAAGAG 1331
769  .....LeuGlnGlyPheGlnGlu 774
1332  GAT 1334
775  Asp 775
seq_name: SwissProt_40:BRFL_HUMAN
seq_documentation_block:
ID BRFL_HUMAN STANDARD: PRT: 1214 AA.
AC P55201: Q9UHI0;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Peregrin (Bromodomain and PHD finger-containing protein 1) (BR140
DE protein).

```

```

GN BRPFL OR BR140.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94161726; PubMed=7906940;
RA Thompson K.A., Wang B., Argaves W.S., Glancotti F.G., Schranck D.P.,
RA Rosolahl E.;
RT "BR140, a novel zinc-finger protein with homology to the TAF250
RT subunit of TFIID.";
RL Biochem. Biophys. Res. Commun. 198;1143-1152(1994).
RP
RP SEQUENCE FROM N.A.
RA Hu S.N., Dong W., Zeng Y.X., Yu J., Yang H.M.;
RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: UNKNOWN. POSSIBLE TRANSCRIPTION ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN TESTIS.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 1 PWM DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M91585; AB02119.1; -.
DR EMBL; AF176815; AAF19605.1; -.
DR MIM; 602410; -.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR000313; PWM.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00855; PWM; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00293; PWM; 1.
DR SMART; SM00355; Znf-C2H2; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50812; PWM; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Zinc-finger; Bromodomain.
FT ZN_FING 21 47 C2H2-TYPE.
FT ZN_FING 273 323 PHD-TYPE.
FT ZN_FING 386 400 C4-TYPE.
FT ZN_FING 645 715 BROMODOMAIN.
FT DOMAIN 1085 1168 PWM.
FT CONFLICT 299 299 E -> A (IN REF. 2).
FT CONFLICT 729 729 V -> L (IN REF. 2).
SQ SEQUENCE 1214 AA; 137542 MW; C530CD2F3083A53D CRC64;

```

alignment_scores:

```

Quality: 245.50 Length: 347
Ratio: 1.240 Gaps: 12
Percent Similarity: 57.061 Percent Identity: 27.089

```

alignment_block:

```

US-09-687-230-1 x BRFL_HUMAN ..
Align seg 1/1 to: BRFL_HUMAN from: 1 to: 1214

```

```

325 ATGACAAACACAGACACAGAAAGCGGAAAAAGAGAAAAAGAGAGAA. 372
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 ILeThrsnArGLeuThrILeGlnArGLysSerGlnPheMetGlnArGLe 540
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 ... AGCGATTCCAGGGAGAAAAAGGGGAAAAAGAGAGAAAGTAA 418
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 uHlSerTyrTrpThrLeuLysArgGlnSerArgGlnSerGlnValProLeuL 557
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 CGAGCAT. .... AAAAGAAAGCAGATCGAGCCGGCTG 452
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 euArGLeuGlnThrHisLeuGlnSerGlnArGLAsnCyAAspGlnVal 573
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 GAGAAATGAGCAGAA. .... AAAGATTCGACTG 481
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
574 GLyArgAspSerGlnAspLysAsnTrpAlaLeuLysGlnLeuLysSe 590
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 TCAGGCCCTGTGAGATTAGTACCTGCTCGTAGAACCTCTGCACACT 531
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 rTrpGlnArGLeuArGLHisAspLeuGlnArGLAlaArGLeuValGlnL 607
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 CTTTAGCCCAACAGAA. .... 548
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
607 euILeArGLysArGLuLysLeuLysArGLuThrILeLysValGlnGln 623
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 ..... GAAGTGAACACAGACACCCCTTCAGAGACTTTGATCAACT 589
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624 ILeAlaMetGlnMetGlnLeuThrProPheLeuLeuLeuArGLysTh 640
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
590 GATGAGACAAATTGCGAGAAAAAGATCCAAAGTCTTTTCATTCTG 639
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 rLeuGlnGlnLeuGlnGlnLysAspTrpThrGlnILePheSerGlnProV 657
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 TGACTGATTTTATTCCTCGCTGCTACTCCATGATTCATTAAACCCCAATG 689
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
657 aLProLeuSerGlnValProAspTyrLeuAspHisILeLysLysProMet 673
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
690 GATTTTACTACCATGAAGAAAGAAATCAAGAAACATGACTTCAGTCCAT 739
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
674 AspPhePheThrMetLysGlnAsnLeuGlnLalaTyrArGLTyrLeuAspPh 690
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 AGAAGAACTAAAGATTAACCTCAAACTATGATGATTCATCCATGATTT 789
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
690 eaSpAspPheGlnGlnAspPheAsnLeuILeValSerAsnLysLeuLysT 707
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 ACAATTAACACAGAGACATTTATTATAAGCTGCAGAAAGACTGTTGCAC 839
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
707 yTAsnAlaLysAspThrILePheTyrArGLAlaLalaValArGLeuArGLu 723
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 TCAGGAATGAATAATCTTACCGCAGAAAGATTCAGACCTCGAAGCAGAG 889
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
724 GlnGlnGlnLalaValArGLAlaLalaTyrArGLGlnAlaGlnLysMetGln 740
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
890 CATAGACTTCATGGCTGACTGCAGAAATCAAGAAAGCAGAAAGATGAGA 939
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 yILeAspPheGlnThrGlnMetHisILePro. HisSerLeuAlaGlnLysp 756
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
940 CAGACACCTTCACAGAGTGGGAGAGACGAGGCTGCTGCGACAGAGAGAGA 989
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GlnLalaThrHisHisThr. .... GlnAspAlaLalaGlnGlnGlnArGLe 771
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 CAGACACTTGAGATGCCGAAGCAGACACCTTCAGAAAGATCCAGCAGAAAGA 1039
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
771 uValLeuLeuGlnLysGlnLysHisLeuPro. .... 781
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 AAATTAAGAAAGAGAAAGATAGCTTGAAGATTAAGTTAAAGCATTA 1089
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
782 ... ValGlnGlnGlnLeuLysLeuLeu. GlnArGLLeuAspGlnValAla 797
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1090 ATTTAGACAGAGAG. .... CAGAGACAGCTTGACCGCATCGTGAAG 1130
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
797 snAlaSerLysGlnSerValGlnArGLSerArGLAlaLysMetILeLys 813

```

```

1131 GAATCTGGAGGAAAGCTGACCGAGCGCTTGTGACACTGACTCGAATT 1180
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
814 LysGlnMetThrAlaLeuArGLArgLysLeuAlaHis...GlnArGLuTh 829
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1181 TGAAGAAGAAAGAACAGAT. .... GGAACAAGACAGCTTGGACTTCGCC 1224
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
829 rGLyArgAspLysProGlnArGLHisGlnLysProSerSerArgLysSerLeuT 846
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1225 ATTCCTGTGATCCCATTTGTAGAGAGCCAGGCTACTGC 1262
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 hrPro. .... HisProAlaAlaLys 852
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: Swissprot_40:CBP_HUMAN

seq_documentation_block:
ID      CBP_HUMAN          STANDARD;          PRT; 2442 AA.
AC      G92793; O16376; O00147;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      CREB-binding protein.
GN      CREBBP OR CBP.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97385172; PubMed=9238046;
RA      Schubo O.M., Borrow J., Tomek R., ReshmI S., Harden A.,
RA      Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA      Zeleznik-Le N.J.;
RT      "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT      acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL      Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97321049; PubMed=9177780;
RA      Giles R.H., Petrif F., Dauwerse H.G., den Hollander A.I.,
RA      Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA      Doggett N.A., Peters D.J.M., Breuning M.H.;
RT      "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT      of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT      16p13.3.";
RL      Genomics 42:96-144(1997).
RN      [3]
RP      SEQUENCE OF 1-405 FROM N.A.
RX      MEDLINE=96376968; PubMed=8782817;
RA      Borrow J., Stanton V.P., Andresen J.M., Becher R., Belin F.G.,
RA      Chaganti R.S.K., Civin C.I., Distche C., Dube I., Fritschau A.M.,
RA      Horman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT      "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
RT      a putative acetyltransferase to the CREB-binding protein.";
RL      Nat. Genet. 14:33-41(1996).
CC      -1- FUNCTION: MEDIATES CBP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC      PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC      THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC      CAMP-RESPONSIVE GENES.
CC      -1- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC      T(8;16)(P11;P13) INVOLVING CBP AND M2, AND T(11;16)(Q23;P13.3)
CC      INVOLVING CBP AND MLL.
CC      -1- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC      SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC      ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC      AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC      -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC      -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its

```



```

906 .....GACTTGCAGAAACTCGAAGACAGAAAGATGAAACGACCTCA 950
      |||
      |||
1213 ystYcIlysgInleuCYstfThrlleProArgrspAlaIaIyTyrSer 1229
      |||
      |||
951 CAGAGTGGGAGGAGCGAGGCTC.....TGCAGAG 982
      |||
      |||
1230 TYrGInAsnArgrTyrHisPhcYsgIuYsCYsPhetHrgIuIeGInI 1246
      |||
      |||
983 AGAGAGAGAGGACTCGAGATCGGAGACACAGCCTTCAGAGTCCCA 1032
      |||
      |||
1246 yGIuAsnValThrlleuGIyAsp.....AspProS 1256
      |||
      |||
1033 GCAGAGAAATATAAAGAAAGACAAAGATATGCTTGAAGTAAAGTTAA 1082
      |||
      |||
1256 eYcInProGInThrlleSerYIsAspInPhcGIuIuYs...LYs 1271
      |||
      |||
1083 AGCAATATTTAGAG.....AG 1099
      |||
      |||
1272 AsnAspThrlleuAspProGIuProPhcValAspCYsGIuCYsGIuAr 1288
      |||
      |||
1100 AGAGCAGAGCAGCTT..... 1115
      |||
      |||
1288 gIysMetHisGInIleCYsValleuHIsTyrAspIleIleIerProSerG 1305
      |||
      |||
1116 .....GACGCATCGTGAAGGAATCTGA..... 1139
      |||
      |||
1305 IyPhcValCYsAspAsnCysLeuIyLysThrGIyArGrProArGIyGIu 1321
      |||
      |||
1140 .....GGAAGCTGACAGCGCGCT 1159
      |||
      |||
1322 AsnIyPhcSerAlaIyArGIeGInThrlThArGIeIuIyAsnHIsIe 1338
      |||
      |||
1160 TGTGAACAGACAGTGCAGATTTGAAGAAAGAAACGATGAAACAGCA 1209
      |||
      |||
1338 uGIuAsnArgrValAsnIyPhcLeuArgrGIuAsn..... 1350
      |||
      |||
1210 CGTTGGACTTCATCCTGTGATCCATTTGTAGAGACCAAGCTAC 1259
      |||
      |||
1351 .....HisPro.....GIuAlaGIyGIu 1356
      |||
      |||
1260 TGCCTGGTGGAGTGCAGATGACGAGAACTGACGTGAGAGCA 1309
      |||
      |||
1357 ValPhcValArgrValAlaIeSer..... 1365
      |||
      |||
1310 TACTTGCAGAGGTTCAAGAGATTAAGCAAAAGTCACTCCAGTGT 1359
      |||
      |||
1366 .....AspIySTHrValGIuValIyProIyM 1375
      |||
      |||
1360 TA.....TATTGAATTATGGCCCTACAGTCTTTATGCACGCCAT 1400
      |||
      |||
1375 eIyIsSerArgrPhcValAspSerGIyGIuMetSerGIuSerPhcProTyr 1391
      |||
      |||
1401 .....TATGACCCACATTTGCAAAATATCAAGAGATGATCTGTATAT 1447
      |||
      |||
1392 ArgrThrlIyAlaIeuphAlaPhcGIuIleAspGIyAlaIyAspAlCY 1408
      |||
      |||
1448 CTAT.....TCAACCTATGGGAGAGCTTCATCTTCATCTTCAA 1482
      |||
      |||
1408 sPhcPhcGIyMetHisValGIuGIuIyGrIySerAspCYsProProIyA 1425
      |||
      |||
1483 GTGAT.....TTGACATCCATGAGTTTGTG 1508
      |||
      |||
1425 snThrArgrArgrValIyIleSerYIleuAspSerIleHisPhcPhcArgr 1441
      |||
      |||
1509 GCCACGTGC.....CAAGATTATCCGTATGTCATGGCAGATAGTAT 1549
      |||
      |||
1442 ProArgrCYsLeuArgrThrlAlaValIyThHisGIuIleuIleGIyTyrLe 1458
      |||
      |||
1550 ACTGAGATGTTTAAACAAAGCA.....GGCATTTCAGACACCTTAC 1550
      |||
      |||
1458 uGIuIyValIyAlaIyLeuGIyTyrValIThGIyHisIle..... 1471

```

```

1591 AAGAGATGGAGATGTCATTGCTGAAGATGAAAGC..... 1625
      |||
      |||
1472 .....TrrAlaCYsProProSerGIuIyAspAspTyrIlePhc 1484
      |||
      |||
1626 CATATCAGACACTTGACACAGAGAAAGAAATGACAGATTACAGAGT 1675
      |||
      |||
1485 HisCYsHisProAsp.....GInIyStI 1493
      |||
      |||
1676 AGAGCCACAGCGCGCTTGGAC..... 1697
      |||
      |||
1493 eProIySProlYsArgrLeuGIuIuTrrYIyIyLysMetLeuAspIySA 1510
      |||
      |||
1698 .....TCCAGT 1703
      |||
      |||
1510 IapheAlaGIuArGIeIleHisAspTyrIyAspIlePhcIyGIuAla 1526
      |||
      |||
1704 ACTCAGACAGCGCTCATAGCGCTGAAGACGTAAACAAATTTGCGCTTC 1753
      |||
      |||
1527 ThrGIuSPArgrLeuHrSerAlaIySGIuIeuprotYrPhc..... 1540
      |||
      |||
1754 AGTTGAATTTTGTACTGTGAAGAGCTGAATATTCAGAGAAACTTG 1803
      |||
      |||
1541 .....GIuGIyAspPhcTrrProAsnValleuG 1550
      |||
      |||
1804 ATGAGACCAACCAATTGCTCAGGAACTCCAGAAAGCCAGAAATGAAGT 1853
      |||
      |||
1550 IuIuIuSer.....IleIySGIuIeGIuIuGIuGIuGIuArGI 1563
      |||
      |||
seq_name: SwissProt_40:CBP_MOUSE
seq_documentation_block:
ID CBP_MOUSE STANDARD; PRT; 2441 AA.
AC P45481;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94019666; PubMed=8413673;
RA Christia J.C., Kwok R.P.S., Lamb N., Hagiwara M., Montminy M.R.,
RA Goodman R.H.;
RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
RL Nature 365:855-859(1993).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZN-ZINC FINGER.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S66385; AAB28651.1; -
CC TRANSFAC: T01318; -
CC MGD: MGI:1098280; Crebbp.
CC InterPro: IPR001487; Bromodomain.
CC InterPro: IPR003101; KIX.
CC InterPro: IPR000197; TAZ_finger.
CC InterPro: IPR000433; znf_22.

```



```

1081 AAAGCAATTAATTTAGAG..... 1097
1272 ysasnAsprThrLeuAspProGluProPheValAspCysylsGlyCysGly 1288
1098 AGAGACGACGACGACGTT..... 1115
1289 ArgLysMetHisGlnIleCysValLeuHisTyrAspIleTleTProSe 1305
1116 .....GACCGCATCGTGAAGAAATCTGGA..... 1139
1305 rGlyPheValCysAspAsnCysLeuLysThrGlyArgProArgLysG 1322
1140 .....GAAAGCTGACCGCGG 1157
1322 LuAsnLysPheSerAlaLysArgLeuGlnThrThrArgLeuGlnLysHis 1338
1158 CTTCGTGAACAGTCAGTCGCAATTTGAAAGCAAAACAGATGAAACAC 1207
1339 LeuGluAspArgValAsnLysPheLeuAlaGlnAsn..... 1351
1208 GACGTTGGAGACTTCTCCATCTGTGATCCCATTTGTAGAGAGCCAGGCT 1257
1352 .....HisPro.....GluAlaGlyG 1357
1258 ACTGCTGTGAGACGAGCGGATGACACACTGCAAGACTTCAGTCTGAGTIG 1307
1357 LuValPheValAlaGlyAlaLaseSer..... 1366
1308 AATTAATTTCAGAGGGTTTCMAAGAGATAAAGCAAAAGTCATCTCAGT 1357
1367 .....AspLysThrValGlnValLysProG 1375
1358 GTTA.....TATTGATTAATGGCCCTACAGTCTTATGACCGC 1398
1375 yMetLysSerArgPheValAspSerGlyGluMetSerGluSerPhePro 1392
1399 AT.....TATGACGCC 1409
1392 yArgThrLysAlaLeuPheAlaPheGluGlnIleAspGlyValAspVal 1408
1410 ACATTTGCCAAATATACAGCAAGCATTTCTGATTTATC..... 1448
1409 CysPhePheGlyMetHisValGlnAspThrAlaLeuIleAlaProHisG 1425
1449 .....TATTCACCTATGGGAGACCTGATCTTCCAA 1482
1425 nIleGlnLysCysValTyrIleSerTyrLeuAsp..... 1436
1483 GTGATTTGACGATCCATGATTTTGGCCACGTC..... 1517
1437 .....SerIleHisPhePheArgProAlaGlyCysLeuArgThrAlaVal 1450
1518 .....CAAGATTATCCGTA 1531
1451 TyrHisGluIleLeuIleGlyTyrLeuGluTyrValLysLysLeuValTyr 1467
1532 TGTCATGCGCATAGTACTGATGATTTTAAACAAGGAGG..... 1574
1467 rValThrAlaHisIleTleThrAlaCysProProSerGlnGlyAspTyrI 1484
1575 .....CATTCACGACCTTA 1589
1484 LePheHisCysHisProProAspGlnLysIleProLysProLysArgLeu 1500
1590 CAAGAGATGAGATGTCATTCCTCGAAGATGAAAGCCATCTAGACACT 1639
1501 GlnGluTyrTyrLysLysMetLeuAspLysAlaPheAlaGluArgIleI 1517
1640 TGACAGAGAAAGAAATGAGAGCATTAAGAGTAGAGCCACGAGGC 1669
1517 eAsnAspTyrLysAspIle..... 1523

```

```

1690 GTTTGACCTCCAGTACACAGACGGCTCATAGCGCTCAAGACGATACA 1739
1524 ..PheLysGlnAlaAsnGlnLysPheLysThrSerAlaLysGluLeuPro 1539
1740 AATTTGGCGTTCAGTTGAAGTTTTCATCTGAGACAGTGAATATAT 1789
1540 TyrPhe.....GluGlyAspPheT 1546
1790 CCAGAAAGAACTTGATGAGACACCATGCTCAGGAACTCAGAGAG 1839
1546 ProAsnValLeuGlnGluSer.....IleLysGlnLeuGlnG 1560
1840 CCAGAAATGAAGCT.....TGAGCACGACGACC 1868
1560 LuGlnGlnGlnArgLysLysGlnLysSerThrAlaAlaSerGluThrPro 1576
1869 CTTGGGAACATGATCTGTCTTGGGTCCCTCATCAGAGAAATGATCTT 1918
1577 GluGlySerGln.....GlyAspSer..... 1583
1919 GCTGACAAAGTGACCAATTAATCTTAAGAAATGACAGCAAGTACTCA 1968
1584 .....LysAsnAlaLysLysLysAsnAsnL 1592
1969 GGTGATATCGTAACAGTATGAGT...TCGAAAGCAATGGGAGATTTC 2015
1592 yLysThrAsnLysAsnLysSerSerIleSerAlaGlnAlaAsnLysLys 1608
2016 CATTCCTTCCCGCTCATGGAACAACACTT 2045
1609 ProSerMetProAsnValSerAsnAspLeu 1618

seq_name: SwissProt_40:T2D1_DROME
seq_documentation_block:
ID T2D1_DROME STANDARD; PRT; 2068 AA.
AC P51123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIID 230 kDa subunit (TAFII-230)
DE (TAFI250) (TBP-associated factor 230 kDa) (P230).
GN TAF250.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 63-75 AND 540-546.
RX MEDLINE=93279463; PubMed=8504928;
RA Kokubo T., Gong D.-W., Yamashita S., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Drosophila 230-kD TFIID subunit, a functional homolog of the human
RT cell cycle gene product, negatively regulates DNA binding of the TATA
RT box-binding subunit of TFIID."
RL Genes Dev. 7:1033-1046(1993).
CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE IN TFIID ASSEMBLY BY
CC INTERACTING WITH BOTH TBP AND OTHER TAF. AS WELL AS SERVING TO
CC LINK THE CONTROL OF TRANSCRIPTION TO THE CELL CYCLE. ESSENTIAL FOR
CC PROGRESSION OF THE G1 PHASE OF THE CELL CYCLE. POSSESSES DNA-
CC BINDING ACTIVITY. IS A NEGATIVE REGULATOR OF THE TATA BOX-BINDING
CC ACTIVITY OF TBP.
CC -!- SUBUNIT: TFIID IS COMPOSED OF TBP AND A VARIETY OF TBP-ASSOCIATED
CC FACTORS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -!- SIMILARITY: TO HUMAN TAFII-250 (CCG1). SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```



```

1671 ysgluGlnTyrAsnGlySerAspThrArgTyrThrLysPheSerLysLys 1687
|||||
1688 IleuGluTyrAlaGlnThrGln..... 1695
1696 ..... 1705
1705 InteuGluAsnAsnIleAlaLysThrGln.....GluArgAlaArg 1718
1721 AlapProGluPheAspGluAlaTrpGlyAsn.....AspAspTyrArg 1734
1734 snphe.....AspArgGlySerArgAlaSerSerProGlyAsp... 1746
1747 .....AspTyr..... 1748
1749 .....IleAspValGluGlyHisGlyHisAlaSerSerSer 1762
1762 snserIleHisArgSerMetGlyAlaGluAlaGlySerSerHisThrAla 1778
1779 ProAla.....ValArgLysProAlaProProGlyArg 1789
1789 OGlyGluValLysArgGlyArgGlyArgProAlaGlyGlnArgAspProV 1806
1806 .....CGTTGGACTCCAGTACGACAGACAGAGCTCATAGCGCTGAAA 1730
1806 alGluGluAspLeuGlnCysSerThrAspGlu..... 1817
1818 ..... 1822
1818 TGAATATTCACAGAAAGAACTTGATGAGACACAGATTCCTGAGGAGAC 1830
1822 uGluAspPheGln.....GluV 1828
1831 TCCGAGAAAGCCAGATGAAAGCTTTGAGCAGCAGAGCCCTCGGAAACATG 1880
1828 alserGluAspGluAsnAlaLaserIleLeuAspGlnGlyArg 1844
1881 ATC 1883
1845 Ile 1845
seq_name: SwissProt_40:P300_HUMAN
seq_documentation_block:
ID P300_HUMAN STANDARD: PRT: 2414 AA.
AC 009472:
DT 15-JUL-1998 (Rel. 36, Created)

```

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DR EIA-associated protein p300.
GN EP300 OR P300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_Taxid=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=95011587; PubMed=7523245;
RA Eckner R., Ewen M.E., Newsome D., Decaprio J.A.,
RA Lawrence J.B., Livingston D.M.;
RT "Molecular cloning and functional analysis of the adenovirus EIA-
RT associated 300-kD protein (p300) reveals a protein with properties of
RT a transcriptional adaptor."
RL Genes Dev. 8:869-884(1994).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS EIA
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZN-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: U01877; AAA18639.1; -.
DR TRANSFAC: T01427; -.
DR MIM: 602700; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZN_ZF.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF02172; KIX; 1.
DR Pfam: PF02135; ZF-TAZ; 2.
DR Pfam: PF00569; ZF; 1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00291; ZNF_ZF; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS00614; BROMODOMAIN_2; 1.
DR PROSITE: PS01357; ZF_ZF_1; 1.
DR PROSITE: PS01355; ZF_ZF_2; 1.
DR PROSITE: PS01356; ZF_ZF_3; 1.
KW Transcription regulation; Nuclear protein; Bromodomain; Cell cycle;
KW Zinc-finger.
FT DOMAIN 11 17 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1067 1139 BROMODOMAIN.
FT DOMAIN 1572 1818 BINDING REGION FOR EIA ADENOVIRUS.
FT ZN_FING 1664 1707 ZN-TYPE.
FT DOMAIN 797 800 POLY-SER.
FT DOMAIN 1519 1526 POLY-GLU.
FT DOMAIN 2066 2069 POLY-GLN.
FT DOMAIN 2190 2195 POLY-GLN.
SQ SEQUENCE 2414 AA: 264143 MW: 68FF909EE4BD693 CRC64;

```

alignment_scores:

```

Quality: 216.00 Length: 740
Ratio: 0.663 Gaps: 33
Percent Similarity: 44.054 Percent Identity: 21.216

```

alignment_block:

```

US-09-687-230-1 x P300_HUMAN

```

```
Align seg 1/1 to: P300_HUMAN from: 1 to: 2414
17 CCGCCCGGC.....GCGCGCGCCCGCTGCCTGCGCGCGG 54
|||||
870 ProProGluProGlnSerGlnAlaLeuHisProProPro..... 882
55 GGTCTCGGGGGCCCGCTCCGCGCTCCGCGTGGCGCGGACGGGA 104
|||||
883 ...ArgGlnThrProThrProThrProThrGlnLeuProGlnGlnValG 898
105 GCGGCGCGCGCACGCGCTGGCGCTGGCGGGGGCGCGGACCGGCGCG 154
|||
898 InProSerLeuProAlaIaIaProSerAlaAspGlnProGlnGlnInPro 914
155 GTCGGAACATGGGCGAAGAACACAAAGACCAAGTC..... 190
|||||
915 ...ArgSerGlnGlnSerThrAlaAlaSerValProThrProAsnAl 929
191 .....GGACAAACACCTTACGA.....GGAGTAT 215
|||||
929 aProLeuLeuProProGlnProAlaIaThrProLeuSerGlnProAlaValS 946
216 GTAGAGAACCCCTTCAAGCTGCTCCTCAAAAGTAGAGAGGAAAGATC 265
|||||
946 eTlIeGlnGlnValSerAsnProProSer.....ThiSerSer 959
266 CGAAGCTCTCAGCGGCGCTCGGCGGACGACCTTCGAAAGACA 315
|||||
960 ThrGlnValAsnSerGlnAlaIaIaGlnGlnGlnPro..... 972
316 AAAGCATCATGACAAACAAAGACGAGCAAGCGGAAAAAGAAAGAAA 365
|||
973 .....SerGlnGlnValValSerMetGlnAlaValSerMetGlnVal 984
366 GGAAGAACACAGATTCACGGGAGAAAGAGGAGAAACGAGAAAGACT 415
|||||
984 aLaspGlnProGlnProAlaAspThrGlnProGlnAspLysSerGlnSer 1000
416 TAAGCAGATTAATAAAGACGAGATCGAGACCGGATGGAATGAGCAG 465
|||
1001 ...LysValGlnAspPcysPylSerMetGlnSerThrGlnThrGln...GluArgS 1016
466 AAAAGATTCACAGTGTACAGCCCGCTGAGATTAAGCTTGCCTCTGAG 515
|||
1016 eTThrGlnLeuLeuSerThrGlnIleLeuSerGlnGlnGlnLysGlnProSerThr 1032
516 AACCTCTCACAAGACTCTTTAGCC.....AAACAAGAAAGAACT 553
|||||
1033 SerAlaThrGlnSerSerProAlaIaProGlnGlnSerLysLysLysIlePh 1049
554 AGAACAGACACCCCTTCAAGAACCTTTGAATCAACTGATGAGACAAATGC 603
|||
1049 eLysProGlnGlnLysLeuArgGlnAlaLeuMetProThrLeuGlnAlaLeu 1066
604 AGAGAAAAGATCCAAAGTCTTC...TTTTCATTTCTGACTGATTTT 650
|||||
1066 yTArgGlnAspProGlnSerLeuProLeuProPheArgGlnProValAspProGln 1082
651 ATT.....GCTCTGGTACTCATCATGATCAATAACACCAATGATTT 694
|||||
1083 LeuLeuGlnGlnIleProAspPylrPheAspIleValLysSerProMetLysPle 1099
695 TACTACCATGAAGAAAGAAAGATCAAGACAAATGACTATCAAGTCCATGAAG 744
|||||
1099 uSerThrIleLeuArgLysLeuAspThrGlnGlnIleGlnGlnProThrG 1116
745 AACTAAAGGTAACCTCAACTATGTGTACTATGCCATGATTTCAAT 794
|||
1116 InTyTValAspAspIleTyrPheLeuMetPheAsnAsnAlaTyrPleuTyAsn 1132
795 AAACAGAGACCATTTATTAAGGCGAAAGACAGCTGTTGCACACAG 844
|||
1133 ArgLysThrSerArgValTyrLysTyrCysSerLysLeu..... 1145
845 AATGAAATTTCTAGCCAGAAAGAAATTCAGAGCTTCAAGACATAG 894
|||||
1146 ...SerGlnValPheGlnGlnGln...IleAspProValMetGlnSerLeuG 1161
895 ACTTCATGCGT..... 905
1161 LyTyTyrCysGlyArgLysLeuGlnPheSerProGlnThrLeuCysCys 1177
906 ...GACTTCAGAGAAACTCGAAAGACAGATGAGAACACACACCTCACA 952
1178 TyTyrGlnGlnLeuGlnCysThrIleProArgAspAlaThrTyrTyrSerTy 1194
953 GAGTGGGAGGAGCAGAGGCTGC.....TGCGACAGAG 984
|||||
1194 rGlnAsnArgTyrThrPheCysGlnLysCysPheAsnGlnIleGlnGlnG 1211
985 AGAGAGAGAGACTCTGGAGATGCCAGACACAGCGCTTCAGAGTCCAGC 1034
|||
1211 uSerValSerLeuGlnAspAspProSerGlnProGlnThrThrIleAsn 1227
1035 AAAGAAAT.....AAAGAAAGACAAAGATATGCTTGAAGATTAAGTTT.. 1079
|||||
1228 LysGlnGlnPheSerLysArgLysAsnAspThrLeuAspProGlnLeuPh 1244
1080 ...AAAAGCAATTAATTAGAGACGAGCAGAGCAGCTT..... 1115
1244 eValGlnCysThrGlnCysGlyArgLysMetHisGlnIleCysValLeuH 1261
1116 .....GACCGCATGCTGAG 1130
1261 lSHsGlnIleIleTyrProAlaGlnPheValCysAspGlyCysLeuLys 1277
1131 GAATCTGGA..... 1139
1278 LysSerAlaIaArgThrArgLysGlnAsnLysPheSerAlaLysArgLeuP 1294
1140 .....GGAAGCTGACGAGCGGCTGTGAAACGTACGTACG 1176
1294 oSerThrArgLeuGlnGlyThrPheLeuGlnAsnArgValAsnAspPheLeuA 1311
1177 AATTGAAAGAAAGAAACACAGATGAGACAGCAGCTTGGCATTCAT 1226
|||
1311 rGArgGlnAsnHisProGlnSerGlnGlnValThrValArgValValHis 1327
1227 CCTGTGATCCCATTTGAGAGAGCCAGGCTACGCTCGTGAGACTGG 1276
|||
1328 AlaSerAspLysThrValGln.....ValLysProGln 1338
1277 AATGAACTGGAAGACTTCAAGTCTGAGTGAATACTTTCGAGGGGTCA 1326
|||
1338 yMetLysAlaIaGlnPheValAspSerGly...GlnMetAlaGlnSerPhe. 1353
1327 AAGAGGATTAAGAAAGAAAGATCACTCCAGTGTATATTGAAATTATGGG 1376
1353 ..... 1353
1377 CCTTACAGTTCTTATGACCCGATTATGACTCCACATTTGCAAAATAG 1426
|||||
1354 ProTyTArgThrLysAla.....LeuPheAlaPheGlnGln 1365
1427 CAAGATGATTTCTGATTTAATCTAT.....TCACCTATG 1461
|||
1365 uLleAspGlnValAspPheCysPhePheGlnMetHisValGlnGlnTyT 1382
1462 GGAAGACTCTGATCTCCAACTGAT..... 1487
1382 LysSerAspCysProProProAsnGlnArgArgValTyrIleSerTyLeu 1398
1488 TTCAGCATCAGATGTTTGGCCACAGTGC.....CAAGATTATCC 1528
|||
1399 AspSerValHisPhePheArgProLysCysLeuArgThrAlaValTyT 1415
```



```

242 uleuArgLeuPheLeuArgAspValThrLysArgLeuAlaThrAspLysA 259
616 CAAAGTCTTTTTCATTTCTGTGACTGATTTTATGTCSTGCTGCTAC 665
259 rgrPheAsnIlePheSerLysProValSerAspTyrLeuGlu..... 272
666 TTCATGATCATTAACACCCCAATGATTTTGTACCATGAAAGAAAGAT 715
273 .....ValIleLysGluProMetLAspLeuSerThrValIleThrLysI 287
716 CAAAGCAATGACTATCATGTCATAGAAGACTAAAGATTAATCTTCAAC 765
287 eAspLysHisAsnTyrLeuThrAlaLysAspPheLeuLysAspLysP 304
766 TAATGTGCTAATGCTCATGATTTACAAATAAACAGAACCATTTATAT 815
304 euIleCysSerAsnAlaLeuGluTyrAsnProAspLys..... 316
816 AAAGCTGCAAGAGCTGTGCACTCAGCAATGAATAATTTCTTACCCAGA 865
317 .....AspProGlyAspLysIleIle.....ArgH1 325
866 AAGATTCAGAGCCTGAAGCAGACATA..... 893
325 sArgAlaCysThrLeuLysAspThrAlaHisAlaIleIleAlaIaGlu 342
894 .....GACTTCATGGCTGACTTCAGAAAACTCGAAAGCAGAAAGAT 935
342 euAspProGluPheAsnLys..LeuCysGluGluIleLysGluAlaArg 358
936 GGAACAG.....ACACCTCACAGATCG 958
358 euLysArgGlyLeuSerValThrSerGluGlnIleAsnProHisSerThr 375
959 GG.....AGGAGGAGGCTGCTGCGCAGAGAGAGAGAG 990
375 LysAlaArgLysThrGluThrArgValGluLysAlaPheArgHisLys 391
991 AGGAGCTGTGAGATGCCGAAAGCAGACCCCTTCAGAGAGTCCCAAGAA 1040
392 ArgAsnProMetLAspValThrHisAsnSerAlaHisLysCysAlaPhe 408
1041 AATTAAGAAAGAGACA.....AAGATATGCTTGA 1069
408 gValArgArgLysSerArgArgArgSerGlnTyrGlyLysGlyIleIle 425
1070 AGATTAAGTTAAAGCAATAATTAGAGAGAGACAGACAGACGCTT.... 1115
425 ys..LysArgGlyValAsnAsnLeuLysLysAspGluGluAspThrLys 441
1116 .....GACCGCATCGTGAAGGAATCTGGA 1139
441 heAlaAspTyrGluAsnHisThrGluAspArgLysIleLeuGluAsnGly 457
1140 GGAAGCTGACACAGCGCTTGTGAACAGTCAGTGCAGATTTTGAAGAG 1189
458 .....GluPheGluValSerThrAspCysHisGluGluAsnGly 470
1190 AAAACAGATGGAACAAGACGTTGACATCTTCATCTGTGGATCCA 1239
470 yGluGluThrGlyAspLeuSerMet..... 478
1240 TTGTAGAGAGACCGGCTACTGCTGTGAGACTGGAGATGACAACTGGA 1289
479 .....ThrAsnAspGluSerSerCysAspLysLe..MetAspLeuAspGlnGly 493
1290 ..AGACTTCAGTCTGAGTGAATACTTTTCAGGGGTTCAAAGAGATAA 1336
494 GluArgLeuAsnAsnGlyAlaGlyThrLysGluAsnPheAlaSerThr 510
1337 AAGGACAAAGTCACTCAGCTGTTATATTTGAATTATGCGCCCTACAGTT 1386
510 uGluGluSerSerAsnGluSerLeuLeuValAsnSerSer.....Sers 525

```

```

1387 CTATGACCGCATATATGACTCCATTTGCAAAATATACAGAAAGATGAT 1436
525 euLeuAsnProGluGlnThrSerArg..... 533
1437 TCGATTTAATCATTTTCACCTATGGGGAAGACTGTGATCTCCAAAGTGA 1486
534 .....LysGluThrPheLeuLysGlyAsnCysLysAsnGly 546
1487 TTCAGCATCCATGAGTTT.....TTGGCCAGCTGGCAAGATT 1524
546 uAlaSerThrAspSerPheGluGlyIleProValLeuGluCysGln.... 561
1525 ATCCGATATGCATGCGCAGATAGTTTACTGATGTTTAAACAAAGAGGCG 1574
562 .....AsnGlyLysLeuGluValValSerPheCysAsp 572
1575 CATTCACGAGCCTACACAGAGATGAGATGTCTATGCTGAGATGAGAG 1624
573 SerGlyAspLysCysSerSerGluGlnLysIleLeuLeuGluAspGln 589
1625 C.....CATACTAGACACTTGACACAGGAAAGAAATGAGAC 1662
589 rLysGluLysProGluThrSerThrGluAsnHisGlyAspAspLeuGlu 606
1663 AGATTACAGAAAGTAGAGCCACAGGCGTTTGACATCTCAGTCAAGAC 1712
606 yLeuGlu.....AlaLeuGluCysSerAsnAsnGlu 616
1713 AGGCTCATGACCGTGAAAGCAGTAAACAATTTTGGCGTTCCAGTTGAGT 1762
617 LysLeu.....GluProGlySerAspValGluVal 626
1763 TTTTGACTGTGAAGAGCTGAATAATTTCCAGAAAGAACTGATGAGACCA 1812
626 LysAspAlaGluLeuAspLysGluGlyAlaSerLysValLysLysTyrA 643
1813 CCAAGATTGCTCAGGGAAGCTCCAGAAAGCCCAAGTGAACGTTTGACACC 1862
643 rGlyLysLeuIle.....LeuGluGlnAlaLysThrThrSerLeuGluLeu 657
1863 AGAGCC 1868
658 ValPro 659
seq_name: SwissProt_40:SPT7_YEAST
seq_documentation_block:
ID SPT7_YEAST STANDARD; PRT; 1332 AA.
AC P35177.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcriptional activator SPT7.
GN SPT7 OR YBR081C OR YBR0739.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
RT important for transcription in vivo.";
RL Genetics 139:523-536(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart O.J.M., Barthé C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;

```



```

640 rva1h1sg1uasnvaa1asnlysncl1uilelysg1uasng1ylysaasng 657
1099 GAGAGCAGAGAGAGCTTGACCCGATCGTAGAGATCTGG.....AGG 1141
657 lnc1ug1uasmetval1g1ug1userSerlyethnc1uaspserserlyys 673
1142 AAGCTGACCGCGCGCTTGACAGCTGACGTCGCAATTGTGAAGAGAA 1191
674 aspdlaasplalal1alylasasprthrg1uaspg1yleu...glnasply 689
1192 AACCAAGATGAAACAGACGTTGGACCTCTCCATCCCTGGATGCCAT 1240
689 srt1ralag1uasenlysg1ualag1ygluasnsng1ug1ug1uaspa 706
1241 .....GTAGGAGCCGACGCTA 1258
706 spaspasp1uaspg1uaspg1uasmetvalaspsertInserlyleu 722
1259 CTGCGCTGTCGACCTGGCAATGACAACTGCACTGCGTGG..... 1303
723 leug1u1yaspaspasp1yaspasp1ueug1u1leservat1yprlysth 739
1304 .....AGTGAATACT 1313
739 rva1thralalysvalargalag1uilegysleuysar1ythr1ylytp 756
1314 TT.....GCAGGGGTTCAAGAGGA 1333
756 helysasncl1ylysa1eua1ns1er1as1er1g1ual1a1p1he1u1y1sa1n1r 772
1334 TAAAGCAACAAGTTCACCTCCAGTGTATATTTGAATATGGCCCCCTCA 1383
772 ocln1r1yeth1y1sar1y1p1he1as1p1g1n1le1u1e1u1y1y1r1y1sg1u1n1l 789
1384 GTTCTTATGACCCGCAATTATGACCTCCAAAT..... 1421
789 ys1ale1ug1u1ser1y1r1a1rg1n1y1s1leg1u1g1n1a1ns1er1le1u1e1t1y1s 805
1422 .....ATGACAGAGATGATTTGTGATTATC..... 1448
806 asng1y1p1he1g1y1th1r1val1leu1y1sg1n1cl1uas1pasp1g1n1leu1g1n1p 822
1449 .....TATTCACCTATGGGAGAGACTCT.....GATCTTCCA 1482
822 eh1as1n1as1p1h1s1er1leu1as1ng1y1a1s1ng1u1a1a1p1he1g1u1y1s1g1n1p1r1o1a 839
1483 GTGATTTACGATCCATGAG.....TTTTCGGCCACGTCGCCAAGATTAT 1536
839 snaspl1eg1u1leu1as1p1r1h1a1rg1p1he1u1.....g1n1g1u1t1y1r 852
1537 CCGTATGTCATGGCAGATGATTCTGATGTTTAAACAAAGAGAGGCA 1576
853 asp.....11es1er1as1n1a1l1e1r1o1s1p1l1e1val1y1y1t1g1u1g1y1a1as 867
1577 TTCAGAGACCTTACAGAGATGAGATCTCATTCCTGAGATGAGAGCC 1626
867 n1th1r1y1s1th1r1e1u1as1p1y1s1met1.....g1n1s1p1l1a1s1er1y 879
1627 ATACAGAGACCTTGACACAGAGAAAGAAATGAGAGATTTACAGAACTA 1676
879 a1as1p1a1rg1met1e1u1g1n1a1s1ng1y1..... 886
1677 GAGCCAGCAGGCGTTGGACTCCAGTACTCAAGACAGGCTCATAGGCT 1726
887 .....11e1as1n1y1s1g1n1s1er1a1rg1p1he1u1a1a1as 896
1727 GAAAGCACTAACAAATTTGGCGTTCCAGTTGAAGTTTGGACTCTGAG 1776
896 n1y1s1.....a1s1p1e1u1g1l1e1u1n1r1p1r1o1y1s1met1.....a1s1ng1n1a 908
1777 AAGCGAAATATTTCCAG.....AAGAAAC1y1t1g1a1t1g1a1c1a1c1a1c1a1g1t1g 1820
:: .....111

```

```

908 sn1le1th1r1e1u1le1g1n1cl1e1a1r1h1s1l1e1c1y1s1h1s1l1e1s1er1leu 924
1821 CTCAGGGAATCTCCAG.....GAAGCCCAAGATGACCT 1853
925 11e1a1rg1met1e1u1s1er1p1r1o1e1s1er1a1l1a1g1n1s1er1a1rg 938
seq_name: SwissProt_40: ATRX_CAEEL

```

```

seq_documentation_block:
ID ATRX_CAEEL STANDARD; PRT; 1359 AA.
AC G9U7E0; C002061.
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Vallard L., Fontes M., Ewbank J.J.,
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RL the human XNP/ATR-X gene."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Fulton R., Wohlmann P.;
RL Submitted (Jun-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF134186; AAD53361.1; -
DR EMBL; AF000196; AAC24256.1; -
DR WormPep; B0041.7; CE17314.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT NP_BIND 496 503
FT SITE 636 639
FT DOMAIN 67 70
FT DOMAIN 266 272
FT DOMAIN 276 281
FT DOMAIN 372 375
FT DOMAIN 603 608
FT DOMAIN 859 862
FT CONFLICT 479 479 C -> F (in Ref. 2).
SQ SEQUENCE 1359 AA; 156191 MW; BBA4342547D4FAE64 CRC64;

```

```

alignment_scores:
Quality: 198.50 Length: 660
Ratio: 0.577 Gaps: 23
Percent Similarity: 52.121 Percent Identity: 21.061

```

alignment block:

US-09-687-230-1 x ATRX_CAEEL ..

Align seg 1/1 to: ATRX_CAEEL from: 1 to: 1359

```
129 GCGCGCGGGGGCGGCGACGGGGCGGCTCGGACATGCGGCAAGAAACACAA 178
|||||
86 AAlaySerSerGluSerGluSerAspGluGluAspArgly 102
|||||
179 GAAACAAAGTCGACAAACACCTCTACAGAGATGATGAGAAAGCCCT 228
|||||
102 eLySerLySerLySlySlyValAspGluLySlySlyGluLySerL 119
|||||
229 TCAACTGCTCTCAAGTAGAGAGGAAGCAACGACGAACTCCACG 278
|||||
119 yLyS.....LysArgThrThr 124
|||||
279 GGCACCTCGGGCGACGACCTCAGCCTTCGAAAGACAAAGATCATGA 328
|||||
125 SerSerSer.....GluAspGluAspSerAspGlu 134
|||||
329 CAACACAAAGACAGACGAAAGGAAAGAAAGAAAGAGAAAGCCCA 378
|||||
134 uGluArgGluGluLySerLySlySerLySlyThrLySlyGluT 151
|||||
379 TTCACGGGGA.....GAAAGGGGAGAAAGGAGAAAGATTAAGAG 422
|||||
151 hSerSerLysSerGluGluSerGluGluGluArgLySlySly 167
|||||
423 GATAAAAAGACGAGATCGACAC.....CGGCTGCAATGAGGC 463
|||||
168 SerLySlyAsnLySlySerValLySlyArgGluGluThrSerL 184
|||||
464 AGAAAAAGTCCCACTGTCAGGCCCTGTGATTAAGATTGCTCCG 513
|||||
184 uGluSerAsp.....GluAspGluLySlyProS 193
|||||
514 AGAAGCCTTCACAAAGCTTTAGCCCAACAGAAAGATAGACAGACA 563
|||||
193 eLyLySerSerLySlyGlyLeuLySlySlyAlaLySerGluSerGlu 209
|||||
564 CCCCTTCAGAAAGCTTGAATCAACGATGAGACATTCAGAGAAAGA 613
|||||
210 SerGluSerLysAspGluLySlyGlu..ValLySlySerLySlyS 225
|||||
614 TCCAGTGCCTTCTTTCATTTCTGCTGACTGATTTTATGCTCGCT 663
|||||
225 rLySlyValValLySlyGluSerGluSerGluAspGluAlaProGlu 242
|||||
664 ACTCCATGATCATTAACACCCCAATGATTTTGTACCATGAAGAAAG 713
|||||
242 yLySlyThrGluLySlyArgLySerLySlyThrSerGluGluSer 258
|||||
714 ATCAAGAAACATGACTATCATGATCAGTAAGAACTAAAGATTAAC 763
|||||
259 SerGluSerLysSerAspGluGluGluGluLySlyGluSerSer 275
|||||
764 ACTAATGTGACTAATGCCATGATTACAATAACAGACACATTTATT 813
|||||
275 o.....LysProLySlySlySlyS 282
|||||
814 ATAAAGCTGCAAGAAAGCTGTG.....CACTCAGA 845
|||||
282 rLeuAlaValLySlyLeuSerSerAspGluLySlyGluGluSerAs 298
|||||
846 ATGAAATTTTGAACGAGAAAGATTCAGAGCCTGAAGACAGACATA 895
|||||
299 ValGluValLeuProGluLySlySlyArgGlyAlaValThrLeuIle 315
|||||
896 CTTTCATGCTGACTTCGAGAAAGCTCAAGAAAGAGATGAGAAAGACA 945
|||||
315 rAspSerGluAspGluLySlyAspGluLySlySerGluSerGluAlaSer 332
|||||
```

```
946 CPTCACAGAGTGGGAGACGAGCGAGCTGTGCGAGAGAGAGAGAGAC 995
|||||
332 alGluGluLySlyValSerLySlySlyAlaLySlyGluGluSerGlu 348
|||||
996 TCTGAGATGCCGAGACACACGCTTCAAGAGTCCGCAAGAAATAATA 1045
|||||
349 SerLySerSerSerGluGluSerLThrValAsnArgLySerLy 365
|||||
1046 AAGAAAGACAAAGATATGCTTGAAGATAG.....TTTAAAGCA 1086
|||||
365 sLySlyGluLySlyProGluLySlySlySlyGlyIleIleMetLys 382
|||||
1087 ATAAATTAAGAGAGACGAGACGACCTTGACCGCATCTGAAGAAAT 1136
|||||
382 eLySlyGluGluGluThrIleAspAlaGluArgGluGluGluArg 398
|||||
1137 GGAGAAAGCTGACACGCGGCTGTGACAGTGCATGCGAATTTGAAG 1186
|||||
399 .....ArgLySlyArgLeuGluLySlySlyGluSerGluPhe.... 409
|||||
1187 AAGAAACCAAGATGGAACMACAGAGTTGGGACTTCTCATCTGTGATC 1236
|||||
410 .....AsnG 411
|||||
1237 CCATTGTAGAGACGACGCTACTGCTGCGTAGAGATGGGAATGCAACT 1286
|||||
411 yLyIleValLeuGluGluGluGluLySlySlySlyThrGlu....MetLeuThr 425
|||||
1287 GGA.....AGACTTCAGTCTGAGTGAATTAATCTTTCAGAGG 1321
|||||
426 GlyThrSerSerGluArgLyLeuLySlySerValIleLeuAspPro 442
|||||
1322 GTTCAAGAGATTAAGAAAGCAAGTACCTCCAGTGTATTAATTTGAT 1370
|||||
442 rSerThrValAspGluGluSerLySlySlyProValGluValHisAsn 458
|||||
1371 .....TATGGCCCTACAGTCTTATGACACCGCAT..... 1400
|||||
458 eLyLeuValArgIleLeuLySlyProHisGluAlaHisGlyIleGluPheMet 474
|||||
1401 TATGACTCCACATTTGCAAAATTCAGCAAGATGATTTGATTAATGTA 1450
|||||
475 TyrAspCysAlaCysGluSerLeuAspArgLeuAspThrGluLySerGlu 491
|||||
1451 TTCACACCTATGGGAAAGCTGATCTCCAGATTCAGATTCAGATCCAT 1500
|||||
491 yGlyIleLeuAlaHisCysMetGlyLeuGlyLySlyThrLeuGluValIle 508
|||||
1501 ACTTTTGGCCACGTCGCAAGATTAATCCGATATGTC.....ATGGCAGAT 1544
|||||
508 hPheLeuHisThrValLeuMetHisGluLySlyIleGlyLeuLySlySly 524
|||||
1545 ACTTACTAGATGTTTAAACAAAGAGGAGCATTCAGACCTCAAGA 1594
|||||
525 ArgValIleValValProLySlyAsnValIleIleAsnThrPheLySly 541
|||||
1595 GATGAGATGTCATTCCTGGAAGATGAAGGCCATTCAGACACTTGACA 1644
|||||
541 uPheGluLySlyThrPheValAspAsnAspGluGluLeuAspThrIleAsp 558
|||||
1645 CAGGAAAGAAATGGAGCAGATTACAGAAATGAGACCAACGCGGTTTG 1694
|||||
558 alAsnGlu.....Leu 561
|||||
1695 GACTCC.....AGTACTCAGACAGGCTCATAGCGCTGAAGACA.... 1733
|||||
562 AspSerLySlyThrIleGluSparGargAlaLeuLySlyAlaThrPhe 578
|||||
1734 .....GTAAACAATTTGGCGTTCCAGTTGAAGTTT 1764
|||||
578 sSerSerLySlyThrProSerValMetIleIleGlyTyrAspLeuPheArg 595
|||||
1765 TTGACTCTGACAGAGCTGAATATTCCAGAAAGAAACTTGATGAGCCACC 1814
|||||
```


FT DOMAIN 1397 1467 BROMODOMAIN 1.
 FT DOMAIN 1520 1590 BROMODOMAIN 2.
 FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
 SO SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;

alignment_scores:

Quality: 181.50 Length: 561
 Ratio: 0.717 Gaps: 21
 Percent Similarity: 45.098 Percent Identity: 21.034

alignment_block:

US-09-687-230-1 x T2D1_HUMAN

Align seg 1/1 to: T2D1_HUMAN from: 1 to: 1872

```

177 AAGAAGCAGACAGTGGACAAACCTCTACGAGAGTATGTAGAGAGCC 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1354 LysLysArgArgValGlyThrThrValHisCysAspTyrLeuAsnArgPr 1370
227 CTGAAAGCTGGTCTCAAGTAGAGGAGAGAGTCAACCGAATCTCCA 276
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1370 HisLysSerLeuHisArgArgArgThrAspPrometValThrLeuSerS 1387
277 CGGCGAGCTCGGGGACGACACTCCAGCTCTTGCAGACAAAGAGATCAT 326
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1387 eTlleLeuGlnSerlleleHisAspMetArgAspLeuProAsnthrTyr 1403
327 GACAAACACAGACAGAAAGCGAAAGAAAGAAAGAAAGAGAGAGACCA 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1404 ProPheHisThrProValAsnAlaLysValValLysAspTyrTyrLysIl 1420
377 GATTCCAGGGGAGAAAGGAGGAGAAAGCGAAGAGAGTAAAGAGATTA 426
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1420 eTlle.....ThrArgPrometAspLeuInThrLeuArgGlnAsnV 1434
427 AAAAGAACCGA.....GATCAGACCGG..... 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1434 aLArgLysArgLeuTyrProSerArgGlnGlnPheArgGlnHisLeuGlu 1450
450 .....GTGGAGATGAGGCGAAGAAAGATCTCCAGTCTACGCCCTGT 493
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1451 LeuIleValLysAsnSerAlaThrTyrAsnGlnProlHisSerLeuTh 1467
494 GAGA.....TTAGACTGCTCGTGGAGAAAGCTCTGCACAA 528
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 gLInLieserGlnSerMetLeuAspLeuCysAspLeuLys..... 1480
529 GCTCTTTAGCCCAACAGACAGAGTAGAACAGACACCCCTTCAAGAGCT 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1481 .....LeuLysGlnLysGlnAspLysLeuAlaArgLeuGlnLysAla 1494
579 TTGAATCACTGATGAGACAAATTCGACAGAAAGATCCAGAGTCTTTC 626
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1495 IleAsnProLeu.....LeuAspAspAspGlnValAlaPheSe 1508
626 ..... 626
1508 rPheIleLeuAspAsnIleValIThrGlnLysMetMetAlaValProAsp 1535
627 .....TTTTCAATTCCTGTGACTGATTTTATGTGCTCTGCTGCTGCT 668
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1535 eTTrProPheHisHisProValAsnLysLysPheValProAspTyrTyr 1541
669 ATGATCATTAACACCCCAATGATTTTATGATCCATGAAAGAAAGATCAA 718
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1542 LysValIleValAsnProMetAspLeuGlnThrIleArgLysAsnIleSe 1558
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
719 GAACAATGACTATCATGCTCATAGAAAGATTAAGATTAATCTCAACTAA 768
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1558 rLysHisIleLysTyrGlnSerArgGlnSerPheLeuAspAspValAsnLeuI 1575
769 TGCTACTAATGCTCATGATTTACAAATTAACCGACAGACCATTTATTA 818

```

```

1575 leleuAlaAsnSerValLysTyrAsnGlnProlGlnSerGlnTyrThrLys 1591
819 GCTGCAAGAGAGCTGTTCCACTCAGAGATGAAATTTCTTACCGAG..... 863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 ThrAlaGlnGlnIleValAsnValCysTyrGlnThrLeuThrGlnTyrAs 1608
864 .GAAGAATTCAGACCTGTAAG..... 884
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 pGlnHisLeuThrGlnLeuGlnLysAspIleCysThrValAlaLysGlnAla 1625
885 .....CAGAGCATGACTTCATGCTGACTGCTGAG 914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1625 lAlaGlnGlnLualGlnLeuGlnLysSerLeuAspPrometThrProGlyPro 1641
915 AAACTGCAAGAGAGAAAGATGAGAAACACCTCCACAGAT..... 956
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 TyrThrProGlnProProAspLeuTyrAspThrAsnthrSerLeuSerMe 1658
957 .GGGAGAGCGAGAGCTGCTGCGAGAGAGAGAGAGAGACTCT.....G 999
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1658 tSerArgAspAlaSerValPheGlnAspGlnSerAsnMetSerValLeuA 1675
1000 GAGATGCCGAGACACAGCCTTCAAGAGTCCACCAAGAAATTAAG 1049
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1675 spIleProSerAlaThrProGlnLysGlnValIThrGlnGlnLysLysP 1691
1050 AAAGCAAAAGAT..... 1061
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1692 GlnAspGlnAspLeuAlaAspGlnGlnLysGlnThrValGlnGlnProG 1708
1062 .....ATGCTTGAAGATTAATTAAGCAATTAATTTAGAGAGAG 1101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1708 nAlaSerValLeuTyrGlnAspLeuLeuMetSer.....GlnLysG 1722
1102 AGCAGAGAGAGCTTGACCGCATCTGTAAGAGATTTGAGAGAAAGCTTACC 1151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1722 LysAspAspGlnGlnAspAlaGlnLysSerAspGlnGlnLysAspAsn..... 1736
1152 AGGGGCTTGTGAACAGTCAAGTGGCAATTTGAAGAAAGAAACCAAGATGG 1201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 .....ProPheSerAlaIleGlnLeuSerGlnLysSerAspSe 1750
1202 AACACAGACGTTGGAGCTTCTCCATCTGTGATCCCATTTGTAGAGAGCC 1251
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1750 rAspValGlnSerGlnLysIleArgProlLysGlnProArgMet..... 1764
1252 CAGGCTACTGCTGTGAGACTGGGATGACAACTGGAACCTTCACTGT 1301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1765 .....LeuGlnGln 1767
1302 GAGATGAATACCTTTGAGGGGTTCAAGAGAGATTAAGAAAGCAAAAGTCA 1351
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 .....Asnthr.....ArgMetAspMetClnAsnGlnLysE 1778
1352 TCCAGTGTATAT..... 1364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1778 rMetMetSerTyrGlnLysArgLysGlnLysAlaLysHisGlnLysGlnA 1795
1365 .....TTGAATTAAGGCGCTTACAGAGTCTTATGACCGCAT..... 1400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1795 spSerAsnIleSerTyrGlnSerTyrGlnLysProAspProlLysSerAsn 1811
1401 .....TATGACTCCACAAATTTGCAAAATTC..... 1424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1812 ThrGlnAspThrSerPheSerIleGlnLysTyrGlnValSerGln 1828
1424 ..... 1424
1828 uGlnGlnAspGlnGlnLysGlnLysGlnArgSerGlnProSerValLeuS 1845
1425 .....AGCAAGAGATTTCTGATTTAATCTATGCA 1454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

678 AACACCCATGATTTAGTACCATGAAAGAAAGATCAAGAAACATGA 727
|||||
331 LysHisProMetAspLeuSerThrValIysArgIysMetGluAsnArgAs 407
728 CTATCAGTCATAGAGAACTAAAGGATATCAACTAATGCTGTCTA 777
|||||
407 PTyArgAspAlaGlnGluPheAlaAlaAspValArgLeuMetHesSerA 424
778 ATGCCATGATTTACATTAACCCAGAACCATTTATTAAGCTGCAAG 827
|||||
424 snCySTyTrpTyrAsnProProAspHisAspValValAlaMetAlaArg 440
828 AAGCTG..... 833
441 LysLeuGlnAspValPheGluPheArgTyrAlaIysMetProAspGluPr 457
834 .TTGCACCTAGCAAGAAATGAATAATCTTACCCAGAAAGAAATTCAGAGCTCGA 882
|||||
457 oLeuGluProGlyProLeuProValSerThrAlaMetProGlyLeu. 473
883 AGCAGACATTAAGACTTCATGCTGACTTGCAGAAAACTCGAAACGAGAAA 932
|||||
474 .....AlaIysSerSer 477
933 GATGGAACAGACACCTCAGAGAGTGGAGAGCGAGGCTGCTGGCAGAG 982
|||||
478 SerGluSerSerGluGluSerSerSerGluSerSerGluGluGlu 494
983 ACAGAGAGAGAGACTCTGAGATGCCGCAAGCACACGCTTCAGAGTCCCA 1032
|||||
494 uGluGluGluAspGluGluAspGluGlu..... 503
1033 GCAAGAGAAATTAAGAAAGAAAGACAAATATGCTTGAAGATTAA 1082
|||||
504 .....Arg 516
1083 AGCAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132
|||||
517 AlaHisArgLeuAlaGluLeuGlnGlnGlnLeu...ArgAlaValHisGlu 532
1133 ATCTGGAGAGAAAGCTGACACAGG 1154
|||||
532 uGlnLeuAlaAlaLeuSerGln 539
seq_name: SwissProt_40:TRDN_HUMAN
seq_documentation_block:
ID TRDN_HUMAN STANDARD: PRT: 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RT and its localisation to chromosome 6q22-6q23.";
RL But. J. Biochem. 233:258-265(1995).
CC -1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE
CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. SARCOPLASMIC
CC RETICULUM.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 018985; AAA75315.1; -.
ID MIM: 603283; -.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 46 BY SIMILARITY.
FT TRANSMEM 47 67 CYTOPLASMIC.
FT DOMAIN 68 728 POTENTIAL.
FT CARBOHYD 74 74 LUMENAL.
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 81423 81423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 728 AA: 81423 MW; C1C53BBE1B20815 CRC64;

```

```

alignment_scores:
Quality: 170.00 Length: 656
Ratio: 0.588 Gaps: 28
Percent Similarity: 44.055 Percent Identity: 20.732

```

alignment_block:

US-09-687-230-1 x TRDN_HUMAN ..

Align seg 1/1 to: TRDN_HUMAN from: 1 to: 728

```

225 CCCTTGAAGCTGCTCTCAAGTAGAGGAGGAAGCAAGTCAACGAA..... 269
|||||
87 ProLeuLysLeuVal...ArgAspAlaMetGluGluThrThrAspTrpI 102
270 .....CTCTCCACGGGACAGCTCGGGGCACAGACTCCA 300
|||||
102 eTyArgIyPhePheSerLeuLeuSerAspIleIleSerGluAspGluG 119
301 GCCTTCTCGAAGCAAAAGCAATCATGACAAA..... 332
|||||
119 LysAspAspAspGlyAspGluAspThrAspThrAspGluIleAspGluPro 135
333 .....CACAAAGACAGAAACCGGAAAGAAAGAAA 361
|||||
136 ProLeuArgLysLysGluIleHisLysAsp...LysThrGluLysGlnI 151
362 GAAAGGAGAGAGAGAGAT.....CCAGGGAAGAAA 393
|||||
151 LysProGluArgLysIleGlnThrLysValThrHisLysGluLysGlu 168
394 AGGGGAGAAAGCGAGAAAGATTAAAGAG.....GATTAAGAAAGAGCA 437
|||||
168 ySGlyLysGluLysValArgGluLysGluLysProGluLysLysAlaThr 184
438 GATCCAGACCGGGGTGAGCAATGAGCGCAAAAGATCTCCAGTGCAGCG 487
|||||
185 HisLysGluLysIleGluLysLys..... 192
488 CCCTGTGAGATTAGACTTCCTCTGAGAGAGCCCTCCACAAAGCTCTTGA 537
|||||
193 .....GluLysProGluThrLysTrpValA 201
538 CCAACACAGAA..... 548
|||||
201 lalysGluGlnLysLysAlaLysThrAlaGluLysSerGluGluLysThr 217
548 ..... 548
218 LysLysGluValLysGlyLysGlnGluLysValLysGlnThrAlaAl 234
549 .....GAAGTAGAAGACAGACACC.....CTTCAAG 573
|||||
234 alysValLysGluValGlnLysThrProSerLysProLysGluLysGluA 251

```

```

574 AACCTTGAATCACTGATGAGACAATTCAGAGAAAGATCCAGTGCCT 623
    ::      ::      ::      ::      ::      ::      |||
251 sPlYsGluLysAlaIaValSerLysHISGluInLysAspInLysAla 267
    |||
624 TTCTTTTCATTCCTGTGACTGATTTTAT 661
    |||      ::      ::      ::      ::      ::      |||
268 pHeCysAlGTYMeTleAspIleAspIleValHISGluAspLeuLysProG1 284
    |||
662 CTACTCCATGATCATTAAACACCCCAATGCATTTTACT 698
    |      |||      |||      ::      ::      ::      |||
284 yGlnSerProAlaIleProProIleuProThrGluInAlaSerArp 301
    699 ..... ACCATGAAGAAAGATCAAGAACAAATGACTAT 731
    301 rOThrProAlaSerProAlaLeuGluLys..... 311
    732 CAGTCCATAGAAAGAACTAAAGATTACTCAATATGTACTAATGC 781
    |||      |||      ::      ::      |||
312 ..... GluGluLysLysAlaGluLysValThrSerG1 325
    782 CATGATTTTCAATTAACAGACATTTATTTAAAGCTGCAAGAACGC 831
    ::      ::      |||      |||      |||      ::      |||
325 uThrLysLysLysGluLysGluAspIleLysLysSerGluLysGlu 341
    832 TGTTCACACTCAGAAATGAATTCCTTAGCCAGAAAGAAATTCAGACCTG 881
    ::      ::      |||      ::      ::      |||      |||
342 ThraIleAspValGluLysLysGluProGluLysAlaSerGlu.Thr 357
    882 AAGCAG..... AGCATGACTTCATGCTGCTGACTTCACAAAACCTG 922
    |||      |||      ::      |||      ::      |||
358 LysGluGluYThrValLysIleAlaIaGluInAlaIaLysAspG1 374
    923 AAGCAGAAAGATGGAGACAGACCTCACAGATGGGAG..... 962
    ::      ::      |||      ::      ::      |||
374 uLysLysGluAspSerLysLysThrLysLysProAlaGluValGluInP 391
    963 .. GACGAGGCTGCTGGCAGAGAGAGAGAGACTCTGAGATGCCGAA 1010
    |||      ::      ::      ::      ::      ::      |||
391 rOlySglY..... LysLysGluGluLysLysGluLysHIS 402
    1011 GCACACGCTTCAGAGTCCCAAGAAAGAAATTAAGAAAGAACCAAA 1060
    ::      |||      |||      |||      ::      ::      |||
403 ValGluProAlaLysSerPro..... LysLysGluInHISe 414
    1061 TATGCTTGAAGATACTTTAAAGCAATTAATTAGAGAGAGACAGAC 1110
    ::      ::      |||      ::      ::      |||      ::      |||
414 rValProSerAspLysGluValLysAlaLysThrGluArgAlaLysGluG 431
    1111 AGCTTAACCGCATGTGAAGAAATCTGA..... GGAAGCTGACCAGG 1154
    ::      ::      ::      ::      ::      ::      |||
431 IuIleGluAlaValSerSerLysLysAlaValProGluLysLysGluIn 447
    1155 CGGCTGTGAAGCATGTCAGTCCGAATTGAAAGAAAGAAACCATGGAAC 1204
    ::      ::      ::      ::      ::      ::      |||
448 LysThrThrLysThr... ValGluGluInGluIleArgLysGluLysSerG1 463
    1205 AAGCAGTTGGACTTCTCCATCCTGTGATCCCATTTAGAGAGACCGAG 1254
    |||      ::      ::      ::      ::      ::      |||
463 yLysThrSerSerIleLeuLysAspLysGluProIleLysGlyLysGluG 480
    1255 GCTACTGCTGCTGTC..... 1268
    ::      ::      ::      ::      ::      ::      |||
480 LuLysValProAlaSerLeuLysGluLysGluProGluThrLysLysAsp 496
    1269 ... AGACTGGGAATGACACTGGAAGACTCAGTGGAGTGAATACTTT 1315
    ::      ::      ::      ::      ::      ::      |||
497 GluLysMetSerLysAlaGlyLysGluValLysProLysProProGluInL 513
    1316 GCAGGGGTTCAAGAGAGATAA..... AGAACAAGATCA 1350
    |||      |||      ::      ::      ::      ::      |||
513 uGluGluLysLysGluGluLysProGluProGluInLLeuLysGluAlaL 530

```

```

1351 CTCACGTGTATATTGATTAATGAGCCCTACAGTCTCTATGACCGCAT 1400
    |||      ::      ::      ::      ::      ::      |||
530 ySProAlaIle..... 533
    1401 TATACATCCACATTTGCAAAATATCCAGACAGATGATTCGATTTAATCTA 1450
    |||      ::      ::      ::      ::      ::      |||
534 ..... SerGluLysValGluIleHISLys..... GluAspIleValLys 546
    1451 TTCACCTATGGGAGACCTGTGATCTTCAGTCAAGTGAATTCAGATCCATG 1500
    ::      ::      ::      ::      ::      ::      |||
546 sProGluLysThrValSerHISGluLysProGluGluLysValLeuLysG 563
    1501 AGTTTTCGCC..... ACGTGCCAAGATTATCCGTATGTC 1535
    ::      |||      |||      ::      ::      |||
563 InValLysAlaValThrIleGluLysThrAlaLysProLysPro..... 577
    1536 ATGCGAGATATTACTGATGTTTAAACAAAGAGGCGCATTCACAGAC 1585
    |||      |||      ::      ::      |||
578 ..... ThrLysLysAlaGluHISArgG1 585
    1586 CCTACAA..... GAGATGAGATGTCATTTGCTGAGAGATGAAG 1623
    ::      ::      ::      ::      ::      |||
585 uArgGluProProSerIleLysThrAspLysProLysProThrProLysG 602
    1624 GCCATCTAGACACTTCACACAGAGAAAGAAATGAGACAGATTAACAGAA 1673
    |||      ::      ::      ::      ::      ::      |||
602 LyThrSerGluValThrGluSerGluLysLysThrGluIleSerGlu 618
    1674 GNAGAGCCACAGGCGTTTGACTTC..... 1700
    |||      ::      |||
619 LysGluSerLysGluLysAlaLysPheLysHISLeuArgGluLysVal 635
    1701 .AGTACTCAAGACAGGCTCATAGCCCTGAAGACAGTAACAAATTTTGCGC 1749
    |||      ::      ::      ::      ::      ::      |||
635 lSerThrArgLysGluSerLeuGluInHISAsnValThrLysAlaGluL 652
    1750 TTCCAGTGAAGTTTTCAGCTGTGAAGAAAGCTGAATATTCAGAGAGAA 1799
    |||      ::      ::      ::      ::      ::      |||
652 ySProAlaArgValSerLysAspValGluAspValProAlaSerLysLys 668
    1800 CTTGATGAGACCACC 1814
    |||      |||
669 AlAlysGluGluYThr 673

```

Sequence	Strd	Score	Len	Documentation
sp_human:09GN11	+	3293.50	4425.16	5.9e-239
sp_human:09U519	+	3287.00	4416.38	1.8e-238
sp_rndent:098665	+	2962.50	3978.96	4.2e-214
sp_human:09B478	+	2283.00	3066.27	4.1e-163
sp_human:096K44	+	1716.50	2305.15	1.3e-120
sp_invertebrate:09VLX2	+	738.50	978.29	4.4e-47
sp_human:09H8M2	+	684.00	909.94	4.8e-43
sp_rndent:09G7M2	+	635.50	854.55	1.7e-39
sp_invertebrate:017581	+	517.50	683.24	1.6e-30
sp_human:09H7R9	+	391.00	522.20	4.1e-21
sp_human:09H5O5	+	274.00	360.31	2.8e-12
sp_human:09Y4Q3	+	255.50	328.94	7.8e-11
sp_human:09U0U1	+	252.00	348.99	8.3e-11
sp_plant:09SM19	+	244.50	311.41	5.5e-10
sp_plant:09S5X1	+	236.00	305.03	2.2e-09
sp_rndent:09Y5V4	+	236.00	302.78	2.2e-09
sp_plant:09JL88	+	229.00	294.09	7.5e-09
sp_human:09S692	+	229.00	292.10	7.8e-09
sp_invertebrate:09V4J4	+	224.00	279.93	2.1e-08
sp_invertebrate:09T068	+	217.50	267.70	7.0e-08
sp_invertebrate:09Q941	+	207.50	256.44	3.7e-07
sp_invertebrate:076561	+	204.00	265.64	5.0e-07
sp_human:09H0E9	+	202.00	251.66	9.1e-07
sp_human:09H3A1	+	201.00	247.97	1.1e-06
sp_human:015355	+	198.00	250.89	1.6e-06
sp_plant:09H7A2	+	196.50	265.16	1.4e-06
sp_human:09H2T5	+	196.00	241.91	2.5e-06
sp_human:09H4T5	+	196.50	241.79	2.5e-06
sp_invertebrate:09J314	+	196.50	241.61	2.5e-06
sp_invertebrate:09V9V3	+	196.50	235.61	2.9e-06
sp_invertebrate:09U753	+	195.00	239.60	3.2e-06
sp_human:09HBD3	+	194.00	238.16	3.2e-06
sp_invertebrate:09U755	+	193.50	237.95	4.2e-06
sp_invertebrate:09O911	+	193.00	244.51	3.9e-06
sp_invertebrate:073897	+	193.00	244.46	3.9e-06
sp_human:09HBD4	+	193.00	236.63	4.6e-06
sp_human:09H3F3	+	192.00	235.44	5.3e-06
sp_fungi:09Y7N0	+	191.50	244.68	4.8e-06
sp_human:04A178	+	191.50	240.29	5.3e-06
sp_human:0969M9	+	191.50	240.29	5.3e-06
sp_human:09NP11	+	651		0.9npl1 homo sapiens (human). b
sp_human:09U519	+	652		0.9nU59 homo sapiens (human). b
sp_rndent:098665	+	651		0.98665 mus musculus (mouse). s
sp_human:09B478	+	651		0.98B48 homo sapiens (human). s
sp_human:096K44	+	651		0.96K44 homo sapiens (human). cc
sp_invertebrate:09VLX2	+	861		0.9VLX2 drosophila melanogaster
sp_human:09H8M2	+	501		0.9H8M2 homo sapiens (human). cc
sp_rndent:09G7M2	+	174		0.9G7M2 mus musculus (mouse). it
sp_invertebrate:017581	+	636		0.17581 caenorhabditis elegans
sp_human:09H7R9	+	233		0.9H7R9 homo sapiens (human). cc
sp_human:09H5O5	+	362		0.9H5O5 homo sapiens (human). cc
sp_human:09Y4Q3	+	715		0.9Y4Q3 homo sapiens (human). cc
sp_plant:09SM19	+	56		0.9npl1 homo sapiens (human). it
sp_plant:09S5X1	+	952		0.9nS19 oryza sativa (rice). sim
sp_rndent:09Y5V4	+	556		0.9S1X1 arabidopsis thaliana (mouse). it
sp_plant:09JL88	+	706		0.9JL88 mus musculus (mouse). it
sp_human:09S692	+	805		0.9S692 homo sapiens (human). d
sp_invertebrate:09V4J4	+	1430		0.9V4J4 drosophila melanogaster
sp_invertebrate:09T068	+	2065		0.9T068 drosophila melanogaster
sp_invertebrate:09Q941	+	1633		0.9Q941 gallus gallus (chicken)
sp_invertebrate:076561	+	374		0.76561 caenorhabditis elegans
sp_human:09H0E9	+	1235		0.9H0E9 homo sapiens (human). b
sp_human:09H3A1	+	1582		0.9H3A1 homo sapiens (human). b
sp_plant:09H7A2	+	757		0.15355 homo sapiens (human). s
sp_human:09H2T5	+	145		0.9H2T5 homo sapiens (human). s
sp_human:09H4T5	+	1602		0.9H4T5 arabidopsis thaliana (mouse). it
sp_invertebrate:09J314	+	1562		0.9J314 homo sapiens (human). b
sp_invertebrate:09U753	+	1634		0.9U753 homo sapiens (human). b
sp_invertebrate:09V9V3	+	3080		0.9V9V3 drosophila melanogaster
sp_invertebrate:09U755	+	1630		0.9U755 gallus gallus (chicken)
sp_invertebrate:09O911	+	1647		0.9O911 homo sapiens (human). s
sp_invertebrate:09B478	+	1568		0.9B478 gallus gallus (chicken)
sp_invertebrate:09O971	+	729		0.9O971 gallus gallus (chicken)
sp_invertebrate:09H7A2	+	733		0.73897 gallus gallus (chicken)
sp_invertebrate:09H8M2	+	1679		0.9H8M2 homo sapiens (human). s
sp_invertebrate:09H3F3	+	1694		0.9H3F3 homo sapiens (human). s
sp_invertebrate:09Y7N0	+	578		0.9Y7N0 schizosaccharomyces pombe
sp_invertebrate:04A178	+	920		0.4A178 homo sapiens (human). th
sp_invertebrate:0969M9	+	920		0.969M9 homo sapiens (human). th

462 GCAGAAAAAGATCTCCAGTGTCCAGCCCTGTGAGATTAGACTGGCTCC 511
 |||||
 101 AAGIuIuYsAspLeuGInCysHisAlaProValArgLeuAspLeuProPr 117
 |||||
 512 TGAGAAAGCTCTCACAGCTCTTTAGCCAAACAGAAAGAAAGTAAAGCA 561
 |||||
 117 oGIuIuYsProLeuThrSerSerLeuAlaIysGInGluValGluGlnT 134
 |||||
 562 CAGCCCTTCAGAGACTTTGAATCACTGATGAGACAAATTGCAGAGAAA 611
 |||||
 134 hProLeuGInCInuAlaLeuAsnGInLeuMetArgGInLeuGInArgYs 150
 |||||
 612 GATCCAGAGCTTTCTTTTCATTTCTGTGACTGATTTTATGCTCCGG 661
 |||||
 151 AsProSerAlaPhePheSerPheProValThrAspPheIleAlaProGl 167
 |||||
 662 CTATCCCATGATCATTTAAACCCCAATGGATTGTATCATGAAGAAA 711
 |||||
 167 yTrSerMetIleIleYsHisProMetAspPheSerThrMetYsGlu 184
 |||||
 712 AGATCAAGACAATGACTATCAGTCCATAGAAACTAAAGGATTAATTC 761
 |||||
 184 yIleIysAsnAsnAspYrGInSerIleGluGluLeuYsAspAsnPh 200
 |||||
 762 AAATATATGTACTATATGCCATGATTTACATTAACCCAGAGACATTTA 811
 |||||
 201 LysIleuMetCysThrAsnAlaMetIleYrAsnYsProGluThrIleY 217
 |||||
 812 TTTTAAAGCTGCAGAAAGACTGTGTGCACACAGAAATGAAGAAATTC 861
 |||||
 217 tYrYrYsAlaAlaYsIleYsLeuLeuHisSerGlyMetYsIleLeuSerG 234
 |||||
 862 AGGAAAGAAATTCAGAGCTCGAAGCAGAGACATAGACTTCATGGCTGG 911
 |||||
 234 InGluArgIleGInSerLeuYsGInSerIleAspPheMetAlaAspLeu 250
 |||||
 912 CAGAAAACCTGGAAGCAGAAAGATGAGACAGACCTCACAGAGTGGGA 961
 |||||
 251 GInuYsThrArgYsGInYsAspGlyThrAspThrSerGInSerIleY 267
 |||||
 962 GAGACGAGGCTGCTGGCAGAGAGAGAGAGACTGTGAGATGCCGAG 1011
 |||||
 267 uAspGlyGlyCysTrpGInArgGluArgGluAspSerGlyAspAlaGlu 284
 |||||
 1012 CACAGGCTTCAGAGATGCCAGCAAGAAATTAAGAAAGCAAGAT 1061
 |||||
 284 IahIslaPheYsSerProSerYsGluAsnYsYsYsAspYsAsp 300
 |||||
 1062 ATGCTTGAAGATTAAGTTTAAAGCAATTAATTAGAGAGAGAGAGAGCA 1111
 |||||
 301 MetLeuGInuAspYsPheYsSerAsnAsnLeuGluArgGluGlnGlu 317
 |||||
 1112 GCTTGAACCGATCGTGAAGAAATCTGAGAGAAAGCTGACCGCGCTG 1161
 |||||
 317 nLeuAspArgIleValYsGInSerGlyGlyYsLeuThrArgArgLeuV 334
 |||||
 1162 TGAACATCTGCTGCAATTTGAAGAAAGAAACCATGGAACAGACAGC 1211
 |||||
 334 aIaAsnSerGInCysGluPheGluArgYsProAspGlyThrThr 350
 |||||
 1212 TTGGAGCTTTCATCTCTGTGATCCCATTTGTAGAGAGAGCCACTAGT 1261
 |||||
 351 LeuGlyLeuLeuHisProValAspProIleValGlyGluProGlyYrCy 367
 |||||
 1262 CCTGTGTGAGACTGGGAATGACACTGGAAGACTTCAGTCTGGAGTGAATA 1311
 |||||
 367 sProValArgLeuGlnCysMetThrArgYsGInSerGlyValAsnT 384
 |||||
 1312 CTTTGGCAGGGGCTCAAGAGGATTAAGAAAGCAAGTCACTCCCGTGTTA 1361
 |||||
 384 hIleuGInGlyPheYsGInuAspYsAsnYsValThrProValLeu 400

1362 TATTGAAATATGGCCCTACAGTTCTATGACACCGCATTATGACTCCAC 1411
 |||||
 401 TyrLeuAsnYrGlyProYrYrSerSerTyrAlaProHisYrYrAspThr 417
 |||||
 1412 ATTTGCAATATATCAGCAAGATGATTCGTGATTAATTCATTCACCTATG 1461
 |||||
 417 rPheAlaAsnIleSerYsAspSerAspLeuIleYrSerThrYrC 434
 |||||
 1462 GGGAAAGACTGTGATCTCCCAAGTCATTCAGCATCCCATGAGTTTGGCC 1511
 |||||
 434 YgluAspSerAspLeuProSerAspPheSerIleHisGluPheLeuAla 450
 |||||
 1512 AGTGCCACAGATTATCCGATATGTCATGCGAAGATAGTTTACTGATGTTT 1561
 |||||
 451 ThrYsGInAspYrProYrValMetAlaAspSerLeuLeuAspValle 467
 |||||
 1562 AACAAAAGAGGGCATTCAGAGACCTACAGAGATGGAATGTGATTCG 1611
 |||||
 467 uThrYsGlyGlyHisSerArgThrLeuGInGluMetGluMetSerLeuP 484
 |||||
 1612 CTGAAGATGAGAGGCCATTAATGAGACTTGACACAGGAAAGAAATGGAG 1661
 |||||
 484 roGluAspGluGlyHisThrArgThrLeuAspThrAlaYsGluMetGlu 500
 |||||
 1662 CAGATTACGAAGTAGAGCCACAGGGCGTTTGGACTCCAGTACTACAGA 1711
 |||||
 501 ...IleThrGluValGluProProGluYrArgLeuAspSerSerThrGInAs 516
 |||||
 1712 CAGGCTCATACGCTGAAAGCAGTAAACAAATTTGGGTTCCATTTGAG 1761
 |||||
 516 pArgLeuIleAlaLeuYsAlaValThrAsnPhGlyValProValGluV 533
 |||||
 1762 TTTTGAAGCTGGAAGAGCTGAATATTCAGAAAGAACTTGATGAGACC 1811
 |||||
 533 alphaAspSerGluGluAlaGluIlePheGInuYsYsLeuAspGluThr 549
 |||||
 1812 ACCAGATTGCTCAGGAACTCCAGAGACCCAGAGATGAACCTTTGAGCAC 1861
 |||||
 550 ThrArgLeuLeuArgGluLeuGInGluAlaGInAsnGInuArgLeuSerTh 566
 |||||
 1862 CAGACCCCTGGAGACATGATCTGCTTGGAGTCCCTATC AGAGAA 1910
 |||||
 566 rArgProProAsnMetIleCysLeuLeuGlyProSerYrArgGluM 583
 |||||
 1911 TGCATCTTGTGTAACAGTACCAATTAATCTTAAGAAAT GCACAGCA 1959
 |||||
 583 eChIleuAlaGluGluValThrAsnAsnLeuYsGluLeuAlaGInGln 599
 |||||
 1960 GTAATCCCAAGTGAATATGTAAGCAGTATGAGACTTCGAAAGCAATGG 2009
 |||||
 600 ValThrProGluYsAlaIleValSerThrYrGlyValArgYsAlaMetG 616
 |||||
 2010 GATTTCATCTCTCCCGCTCATGGAAGAAACACTTTGTGATTTGACAG 2059
 |||||
 616 YlleSerIleProSerProValMetGluAsnAsnPhValAspLeuThrG 633
 |||||
 2060 AAGACATGAACACTTAAAGAGCGATGTTGATGAGTGTGACCTGCT 2109
 |||||
 633 luAspThrGluGluProYsYsThrAspValAlaGluCysGlyProGly 649
 |||||
 2110 GGAAGT 2115
 |||||
 650 GlySer 651

seq_name: sp_human:Q9UH59

seq_documentation_block:

ID Q9UH59 PRELIMINARY: PRT: 652 AA.

AC Q9UH59: 01-MAY-2000 (TREMBlrel. 13. Created)

DT 01-MAY-2000 (TREMBlrel. 13. Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17. Last annotation update)

DE BROMODOMAIN PROTEIN CELTIX1.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stahl A., Enserink J., Stein J.L., Stein G.S., van Wijnen A.J.;
 RT "Molecular analysis of the human bromodomain protein Celtix-1.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF213969; AAF19526.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain.1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO.1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 SQ SEQUENCE 652 AA; 74084 MW; 70F2B654B2618529 CRC64;

alignment_scores:
 Quality: 3292.00 Length: 652
 Ratio: 5.104 Gaps: 0
 Percent Similarity: 98.926 Percent Identity: 98.620

alignment_block:
 US-09-687-230-1 x Q9UH59 ..

Align seg 1/1 to: Q9UH59 from: 1 to: 652

162 ATGGGCAAGAGCACAAGACAGCAAGTCGGACAAACCTCTAGAGAGA 211
 |||||
 1 MetGlyLysHisLysHisLysSerAspLysHisLeuTyrGlu 17
 212 GTATGTAGAGAAAGCCTTGAAGCTGTCTCAAGTAGAGGAGAAAGAG 261
 |||||
 17 uTyValGluLysProLeuLysLeuValLeuLysValGlyLysGlu 34
 262 TCACCGCAACTCTCCACGGGACGCTCCGGGACAGCTCCACCTCTTGAA 311
 |||||
 34 alThrGluLeuSerThrGlySerSerGlyHisAspSerLeuPheGlu 50
 312 GACAAACAGATCATGACAAACAGACAGAGAAAGCGAAAGAGAGAA 361
 |||||
 51 AspLysAsnAspHisAspLysHisLysAspArgLysArgLysGly 67
 362 GAAAGGAGAGACAGATTCACGGGAGAGAAAGCGGAGAAACGAGAA 411
 |||||
 67 sLysGlyLysGluLeuLeuProGlyGluGlyLysGlyArgLysArgThr 84
 412 GAGTTAAGAGCAATAAAGAGAGAGATGAGACCCGCTGAGACATGAG 461
 |||||
 84 rGValLysGluAspLysLysLysArgAspArgAspGlnValGluAsnGlu 100
 462 GCAGAAAGATCTCAGATGTCAGCGCCCTGTGATTAAGATTCCTGCC 511
 |||||
 101 AlGluLysAspLeuGlnCysHisAlaProValArgLeuAspLeuProP 117
 512 TGAGAGCCCTCTCAGAGCTTTTACCCAAAGAGAAAGAGAGAGAGA 561
 |||||
 117 OGluLysProLeuThrSerSerLeuAlaLysGlnGluGluValGluGln 134
 562 CACCCCTTCAGAGAGCTTTGATCACTGATGAGACAAATTCGAGAGAAA 611
 |||||
 134 hrProLeuGlnGluAlaLeuAsnGlnLeuMetArgGlnLeuGlnArgLys 150
 612 GATCCAAAGTCTTCTTTCATTTCTGTGACTGATTTTATGCTCTGCG 661
 |||||
 151 AspProSerAlaPhePheSerPheProValThrAspPheLeuAlaProGlu 167
 662 CTACTCATGATCATTTAAACCAATGATTTTATGACATGAGAGAAA 711
 |||||
 167 yTyrSerMetLeuLeuLysHisProMetAspPheSerThrMetLysGlu 184
 712 AGATCAAGAACAAATGACTATTCATGATCAGAGAGAGAGAGAGAAATTC 761

|||||
 184 yLysLeuAsnAsnAspTyrGlnSerLeuGluLeuLysAspAsnPhe 200
 762 AAATATATGTGTACTATGCGCATGATTACAAATAAACGAGACGCTTTA 811
 |||||
 201 LysLeuMetCysThrAsnAlaMetLeuTyrAsnLysProGluThrLeu 217
 812 TTATTAAGCTGCAAGAGAGCTGTGCACTCAGAGAAAGAAATTTAGCC 861
 |||||
 217 rTyrLysAlaAlaLysLysLeuLeuHisSerGlyMetLysIleLeuSer 234
 862 AGCAAGAAATTCAGAGCTTGAACAGACAGATGACATTCATGCTGACTTG 911
 |||||
 234 GlnLysThrArgLysGlnLysAspGlyThrAspThrSerGlnSerGly 250
 912 CAGAAAGCTGAAAGCAGAAAGATGGAGACAGACCTCCACAGGTGGGG 961
 |||||
 251 GlnLysThrArgLysGlnLysAspGlyThrAspThrSerGlnSerGly 267
 962 GAGAGAGAGCTGCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
 |||||
 267 uAspGlyGlyCysTyrProGlnArgGlnArgLysAspSerGlyAspAlaGlu 284
 1012 CACAGCCTTCAGAGAGTCCAGCAAGAAAGAAATTAAGAAAGAGAGAT 1061
 |||||
 284 IahIsaLysPheLysSerProSerLysGluAsnLysLysAspLysAsp 300
 1062 ATGCTTGAAGATTAAGTTTAAAGCAATTAATTAGAGAGAGAGAGAGCA 1111
 |||||
 301 MetLeuGlnAspLysPheLysSerAsnAsnLeuGlnArgGlnGlnGlu 317
 1112 GCTTGAACCGCATCGTAAGAGAAATCTGAGAGAAAGCTGACCGAGCGCTTG 1161
 |||||
 317 nLeuAspArgIleValLysGlnSerGlyGlyLysLeuThrArgArgLeu 324
 1162 TGAACAAGTCAGTGCAGATTTGAAGAAAGAAACCAATGGAAGAGAGAG 1211
 |||||
 334 alAsnSerGlnCysGluPheGlnArgArgLysProAspGlyThrThrThr 350
 1212 TTGGAGCTTCTCATCCTGATGATCCATGTAGAGAGAGAGAGAGAGAG 1261
 |||||
 351 LeuGlyLeuLeuHisProValAspProIleValGluProGlyTyrCys 367
 1262 CCTGTGAGAGCTGGGAATGACAACTGAGAGACTTCAGCTGAGTGAATA 1311
 |||||
 367 sProValArgLeuGlyMetThrThrGlyArgLeuGlnSerGlyValAsn 384
 1312 CTTTGAAGGGGTTCAAGAGAGATTAAGAAAGAAAGCAAGTCCAGTGTTA 1361
 |||||
 384 hrLeuGlnGlyPheLysGluAspLysArgAsnLysValThrProValLeu 400
 1362 TATTTGAATATGGGCTTACAGTTCATTATGACACCGCATATGATTCAC 1411
 |||||
 401 TyrLeuAsnTyrGlyProTyrSerSerTyrAlaProHisLysTyrAspSer 417
 1412 ATTTGCAAAATATGACAGAGATGATTCGATTAATCTATTAACCTPAG 1461
 |||||
 417 rPheAlaAsnIleSerLysAspAspSerAspLeuIleTyrSerThrTyrG 434
 1462 GGGAGAGCTGTGATCTTCAAGATGATTTTCAGATCCATGATGATTTGGCC 1511
 |||||
 434 LysLysPheSerAspLeuProSerAspPheSerIleHisGluPheLeuAla 450
 1512 ACGTGCAAGATTAATCCGTATGTCATGAGAGAGATTTACTGATGATTT 1561
 |||||
 451 ThrCysGlnAspTyrProTyrValMetAlaAspSerLeuLeuAspValLe 467
 1562 AACAAAGAGAGGCGATTCACGAGCCTTACAGAGAGATGAGATGTCATG 1611
 |||||
 467 uThrLysGlyGlyHisSerArgThrLeuGlnGlnLeuGlnMetSerLeu 484
 1612 CTGAGATGAGAGGCAATGACTAGAGACCTTGACAGAGAAAGAAATGAGG 1661
 |||||

484 rogluaspqluylhsthrargrhluaspptlhalalsglumetclu 500
1662 CAGATTACAGAGTAGAGCCAGCCAGGGCGTTGGACTCCAGTACTCAAGA 1711
501 glnlletlthngluvalgluProProgllyargluaspserSerThrlglns 517
1712 CAGGCTCATAGCGGCGAAAGCAGTAACAAATTTGGCGTTCCAGTTGAAG 1761
517 pargleuilealaleuylsAlaValThrAsnphleglyalProvalgluv 534
1762 TTTTGGAGCTGAGAGAGCTGAATATTCACAGAACTTATAGACC 1811
534 alpnehspergluclunaglunleuInglunlagnlnsnglunlArgluSerTh 567
1812 ACCAGATTGCTCAGGGAATCCAGGAAGCCCAAGATGAACGTTTGAAC 1861
551 ThrArgluLeuAlargluLeuInglunlagnlnsnglunlArgluSerTh 567
1862 CAGACCCCTGGGAACATGATGTCTCTTGGCTCCCTCATC.AGAGAA 1910
567 rArgProProAsnMetIleCysleuLeuIleProSerTyArgglum 584
1911 TGCACTTGTGACACAGTACCAATTAATCTTAAGAAAT.GCACAGCA 1959
584 ehlslenuaglucInvalThrAsnAsnleuylsGlunleuAlaglIn 600
1960 GTAATCCAGTATATGTAAGACGTAATGAGTTCGAAAGCAATGGG 2009
601 ValThrProgllyAspIleValserThrlyglValArgllyAlamegl 617
2010 GATTTCCATTCCTTCCCGCTCATGGAAAAACAATTGTGCGTTTGACAG 2059
617 ylleSerIleProSerProValMetglunsnAsnphValAspLeuThrg 634
2060 AAGACGTGAAGACCTAAAAAGAGGATGTGATGAGTGCGACTGGT 2109
634 lAspThrleuclunProlyslsThrAspValAlaglucylsGlProgly 650
2110 GGAACT 2115
651 glySer 652

seq_name: sp_rodent:088665

seq_documentation_block:
ID 088665 PRELIMINARY; PRT: 651 AA.
AC 088665;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BROMODOMAIN-CONTAINING PROTEIN BP75.
GN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE=99456832; PubMed=10526152;
RA Cuppen E., van Ham M., Peeters B., Wieringa B., Hendriks W.:
RT bromodomain-containing protein.";
RL FEBS Lett. 459:291-298(1999).
MD MGI:1349766; Brd7.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: Pf00439; Bromodomain; 1.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SO SEQUENCE 651 AA; 74000 MW; 5D34B4F14FD51350 CRC64;

alignment_scores:
Quality: 2966.50 Length: 652
Ratio: 4.701 Gaps: 1
Percent Similarity: 96.779 Percent Identity: 87.117

alignment_block:
US-09-687-230-1 x 088665 ..
Align seg 1/1 to: 088665 from: 1 to: 651

162 ATGGGCAAGAAACACAGAGCAAGTCGGCAAAACCTCTACAGGA 211
1 MetGlyLysLysHisLysLysHisLysSerAspArgHisPheTyrluGl 17
212 GTATGTAGAGAAAGCCCTGAAGCTGTCTCTCAAGTACAGAGGAAG 261
17 uTyrlValgluLysProleuLysLeuValleuLysValglYglYsergluv 34
262 TCACCGAACTCTCCACGGGACGCTGGGGACGACTCCAGCCTCTTCA 311
34 AlThrGluleuSerThrlySerSerGlyHisAspSerSerleuPheglu 50
312 GACAAACGATCATGACAAACACAGGACAGAAAGCGAAAGAAAGAAA 361
51 AspArgSerAspHisAspLysHisLysAspArgLysArgLysLysArgly 67
362 GAAAGAGAGAAAGCAAGATTCAGGGGAAAGAAAGGAGAAACGAGAA 411
67 sLysglYgluLysGlnAlaProlglYgluLysglYArgLysArgArg 84
412 CAGTTAAGAGAGATTAAGAAAGAGAGATCGAGACCGGCTGAGATAG 461
84 rValAllysluAspLysLysLysArgAspArgAspArgAlaglunsnclu 100
462 GCAAGAAAGATCTCCAGTGTACGCGCCCTGTGAGATTAGACTGCTCC 511
101 ValAspArgAspLeuGlnCysHisValProIleArgluLeuAspLeuPro 117
512 TGAGAAAGCCTCTCACAGCTCTTTAGCCAAACAGAAAGATAGACAGA 561
117 ogluLysProleuThrSerSerleuAlalyseIngluValgluInl 134
562 CACCCCTTCACAGACCTTTGAATCAACTGATGAGCAATTCAGAGAAA 611
134 hrProleuGlnleuAlaleuIngnleuMetArggluLeuInlArglys 150
612 GATCCAGTCTTCTTCTTCTTCTTCTTCTGAGACTGATTTATTCGCC 661
151 AspProSerAlaPhePheSerPheProValThrAspPheIleAlaProl 167
662 CTACTCATGATCATTAACACCAATGATTTAGTACATGAAGAAA 711
167 TyrlSerMetIleIleLysHisProlMetAspPheSerThrMetLysglu 184
712 AGATCAAGACAACTATCATCATGCTCATAGAGAACTTAAGATACATTC 761
184 yslleLysAsnAsnAspTyrlGlnSerIleGluGluleuLysAspAsnph 200
762 AAATAATGTGTACTAATGCCATGATTTAACAATAACAGAGACCATTTA 811
201 LysleuMetCysThrAsnAlaMetIleTyrlAsnLysProgluThrIleTy 217
812 TTATAAGCTGCAAGAAAGCTGTTCACCTCAGATGAAGAAATTTCAACC 861
217 TyrlLysAlaAlaLysLysleuLeuHisSerclYmetLysIleleuSer 234
862 AGGAAGAAATTTAGAGCCTGAAGCAGAGACATAGATTTATGGCTACTG 911
234 lngluArgIleGlnSerleuLysGlnSerIleAspPheMetSerleu 250
912 CAGAAACCTCGAAAGCAGAAAGATGAACAGACACCTCAGAGAGTGGGGA 961
251 GlnLysThrArgLysGlnLysGlnLysArgThrAspAlaCysGlnSerclYgl 267

53	pheMetAlaAspLeuGlnGlnYstThrArgLysGlnLysAspLysThrAspThr	69
947	CTCACAAGATGGGAGCAGCAGGAGCTGCTGGCAGACAGACAGAGAGACT	996
69	rserGlnserGlnGlnAspArgLysGlyCysTTPGlnArgGlnArgGlnAsp	86
997	CTGGAGATGGCCGAAGCAGCAGCCTTCAAGAGTCCACCAAGAAATATAA	1046
86	ergLysAspAlaGlnAlaHisAlaPheLysSerProSerLysGlnAsnLys	102
1047	AAGAAAGCAAAAGATATGCTTGAAGATAACTTAAAGCAATAATTACA	1096
103	LysLysAspLysAspMetLeuGlnAspLysPheLysSerAsnLysGln	119
1097	GAGAAAGACAGAGCAGCTTACGCCCATCTGAAGAGATCTGGAGGAAGC	1146
119	narGlnGlnGlnGlnLeuAspArgLysLevalLysGlnSerLysGlnLysL	136
1147	TGACCAAGCAGGCTTGTGAACAGTACAGTCGAAATTTGAAGAAGAAACA	1196
136	eutThrArgArgLeuValAsnSerGlnCysGlnPheGlnArgArgLysPro	152
1197	GATGGACACAGCAGCTTGGGACTTCTGCATCTGTGCATCCCATTTGAGG	1246
153	AspGlyThrThrThLeuGlnLysLeuHisIserovalAspProIleValGln	169
1247	AAGAGCAGGCTTCCTCCGAGAGCAGCTGGAAATACAACTGGAGACTTC	1296
169	ylGlnProGlnLysLysProvalArgLeuGlnMethThrThGlnArgLeuG	186
1297	AGCTGGAGTAGATCTTGCAGGGGCTTCAAAAGAGATAAAGAACAAA	1346
186	InserrGlyValAsnThrLeuGlnGlnPheLysGlnAspLysArgAsnLys	202
1347	GNACAGCCAGCTTATTAATTTGAATATGAGCCCTCAAGTTGTATGACC	1396
203	ValThrProvalLeuThrLysAsnLysLysLysProLysSerSerLysAlaIrr	219
1447	TCTATTCAACCTATGGGGAAGACTGTGATCTCCAAAGATTAGATC	1496
236	LeuYserThrLysLysLysAspSerAspLeuProSerAspPheSerLys	252
1497	CATGAGTTTTTGGCCACGTCGCCAAGATTATCCGATGTGCATGCAGATAG	1546
253	HisGlnPheLeuAlaThrCysGlnAspLysProLysValMetAlaAspSe	269
1547	TTTTACTCGATGTTTAAACAAGAGGGCATTCCAGAGCCCTACAAAGGA	1596
269	rLeuLeuAspValLeuThrLysGlnLysLysIserThrGlnLeuGlnGln	286
1597	TGGAGATGTCATTTGCTGTGAAGTGAAGGCCAATCTAGACACTTGACACA	1646
286	etGlnMetSerLeuProGlnAspArgLysHisThrArgThrLeuAspThr	302
1647	GGAAAGAAATGGAGCAGATTCACAGATGAAGGCCACACAGGGCTTTGGA	1696
303	AlaLysGlnMetGlnLysLysLysLysLysLysLysLysLysLysLysLys	319
1697	CTCCAGTACACAGAGAGCCTCATAGCCGTGAAGACAGTAACTTAATTTTG	1746
319	pSerSerThrGlnAspArgLeuThrAlaLeuLysAlaValThrAsnPheG	336
1747	GCCTTCCAGTTGAAGTTTTTGACTGTGAAGAAAGTGAATATTTCCAGAG	1796
336	LysAlaProvalGlnValAlaPheAspSerGlnLysAlaGlnLysLysLys	352
1797	AAACTTGATAGACACACAGATTTGCTCAGGGAACATCCAGAGCCACADA	1846
353	LysLeuAspLysLysThrThrArgLeuLeuAspGlnLeuGlnLysAlaGlnHis	369

```

1847 TGACGCTTTGAGCACCAGACAGCCCCCTGGGAAACGATGATCGCTCTGGGCT 1896
1848 TGAACGCTTTGAGCACCAGACAGCCCCCTGGGAAACGATGATCGCTCTGGGCT 1897
369 ngIarGlsuIeSerThrArgProProProPronsmetIleCysIeuIeuGlyP 386
1897 CCTCATC.AGAGAAATGCATCTTGTCTGAAACAAGTACCAATATCTTAA 1945
1898 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
386 roserTyrArgGluMetHisIeuAlaGluGlnValThrAsnAsnIeuLys 402
1946 GAATT.GCACAGCAAGTAACTCCAGGTGATATTCGTAAGCACTATGAGCT 1994
403 GluIeuAlaGlnGlnValThrProGlyAspIleValSerThrTyrGlyVa 419
1995 TCGAAAAGCAATGAGGGAGTTTCATCTCTTCCCCTCATGAAAACAACT 2044
419 IarGlyAlaMetCylIleSerIleProSerProValMetGluAsnAsnP 436
2045 TTGTGATTGTACAGAAAGACACTGGAAGAACTTAAACAGCATGTTGAT 2094
436 heValAspIeuThrGlnAspThrGluGlnIuproLysIysThrAspValAla 452
2095 GAGTCTGCACTGCTGGTGAAGT 2115
453 GluGlyGlyProGlyGlySer 459
seq_name: sp_human:096KR4
seq_documentation_block:
ID 096KR4 PRELIMINARY: PRT: 351 AA.
AC 096KR4:
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE CDNA FLJ14402 FIS, CLONE HEMBA1003783, MODERATELY SIMILAR TO MUS
DE MUSCULUS BROMODOMAIN-CONTAINING PROTEIN BP75 MRNA.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO, AND MAINLY HEAD;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watsutsumi M., Hosoki T., Raku Y., Kodaira H., Kondo H., Sugawara
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi
RA Watanabe J., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.
RA Yamamoto S., Wakamatsu S., Nakamura Y., Nagahari K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK027308; BA855931.1; - 51FLID2ABD7DEDA0 CRC64;
SQ SEQUENCE 351 AA; 3236 MW; 51FLID2ABD7DEDA0 CRC64;

```


alignment_scores: Quality: 738.50 Length: 824
 Ratio: 1.656 Gaps: 24
 Percent Similarity: 54.126 Percent Identity: 28.034

alignment_block:
 US-09-687-230-1 x 09VLX2 ..

Align seg 1/1 to: 09VLX2 from: 1 to: 861

```

168 AAGAACACAAAGACACAGTCGAC...AACACCTCTACGAGAGTA 214
|||||.....:|||||.....:|||||.....:|||||.....:
5  LysLysHisLysLysAsnLysSerGluArgGluLysTyrGluGluTyr 21
215 T.....:GTAGAAAGCCCTGAGCTGCTCC 240
|.....:.....:.....:.....:.....:.....:
21 rSerGlnHisGlnAspProAlaGlnLeuGlnArgGluLysLeuLysLeu 38
241 TCAAAGTAGAGGAGGAGAGTCCAGCACTCTCCAGGGCAGCTCGGG 290
|||||.....:|||||.....:|||||.....:|||||.....:
38 euLysValGlySerAsnAlaThrProGluTyrSerAlaAsnSerPromet 54
291 CACGAC.....:TCACGCTCTT 307
||.....:.....:.....:.....:.....:.....:
55 ValAspGlyGlyProProThrAlaAlaGluAlaMetMetSerProValPr 71
308 CGAAGACAAAGACATCATGACAAACACAGACAGAAAGCGAAAGA 357
|||||.....:|||||.....:|||||.....:|||||.....:
71 oGluGluLeuGlnAspHisGlnGlyHisArgGluArgHisLysLysSerL 88
358 GAAAGAAAGAGAGAGAGATTCAGGGGAGAAAGGAGAGAAACGG 407
:|||||.....:.....:.....:.....:.....:.....:
88 yLysLysLysLysLys.....:LysAspArgGluLysLysHis 100
408 AGAAGAGTTAAGAGAGATAAAGACGAGATCGAGACCGGTGGAGAA 457
:|||||.....:|||||.....:|||||.....:|||||.....:
101 LysHisHisLysGlu...LysArgHisArgSerArgAspArgHisArgHis 116
458 TCGAGCGAAG.....:AAAGATCTCCACTCTC 483
:.....:.....:.....:.....:.....:.....:
116 pAlaGlySerAspGluAspMetMetAlaGlyAlaAspAspAlaAlaCys 133
484 AC.....: 485
133 erGlyPheAlaProSerSerValAlaProProAlaAlaAspProAspSer 149
485 .....: 485
150 SerGlnAspGlyPheSerPheMetAspAspGlnSerGlnProLeuPr 166
485 .....: 485
166 oGluAsnLleLeuPhePheAlaGlyLysThrThrAspAsnSerProSerA 183
486 .....:GCCCTGTGAGATTAGC.....: 503
|||||.....:|||||.....:|||||.....:
183 snCysProValThrLysProLleAlaProArgLysLeuAspAspLleLeu 199
503 .....: 503
200 MetGlySerSerProAsnSerSerSerLeuGlnSerSerSerLeuGlyLe 216
504 .....:TTGCCT.....: 509
|||||.....:|||||.....:|||||.....:
216 uLleGlySerSerProThrLysProLeuProAspLeuLleProSerP 233
509 .....: 509
233 roSerThrProGlyGlyAlaAsnSerLeuAsnAlaLeuThrProLysAla 249
510 .....:CCTGAGAGCCTCTCACAAAGCTTTAGCCAAACAGAA... 548
|||||.....:|||||.....:.....:.....:

```

```

250 LeuGluAlaProLysThrProSerSerSerSerGluArgGluPr 266
549 .....:GAGTAGAACAGACACCCCTTCAAGAG 576
266 oArgSerCysValLeuLysLysLeuGlnGlnLysSerProLeuAsnLysL 283
577 CTTTGAAATCAACTGATGAGACAAATTGCAGAGAAAAGATCCAGTGCCTTC 626
|||||.....:|||||.....:|||||.....:|||||.....:
283 euLeuGlnHisLeuLeuArgPheLeuGluLysArgAspProHisGlnPhe 299
627 TTTTCATTTCCTCGACATGATTTTATGCTCTGCTACTGACATGATCAT 676
|||||.....:|||||.....:|||||.....:|||||.....:
300 PheAlaThrProValThrAspAspMetAlaProGlyTyrSerSerLleI 316
677 TAAACACCCCAATGATTTTATGACCATGAAGAAAGATCAAGAACAAATG 726
|||||.....:|||||.....:|||||.....:|||||.....:
316 eSerArgPrometAspPheSerThrMetArgGlnLysLleAspAspHisG 333
727 ACTATCAGTCCATAGAGAACTAAAGATTAACCTCAAACTAATGTGTACT 776
|||||.....:|||||.....:|||||.....:|||||.....:
333 LuTyrThrAlaLeuThrGluPheThrAspAspPheLysLeuMetCysGlu 349
777 AATGCCATGATTTACAAATTAACACAGACACATTTTAAAGCTGCAAA 826
|||||.....:|||||.....:|||||.....:|||||.....:
350 AsnAlaLleLysTyrAsnHisValAspHisValItyAsnLysAlaAlaL 366
827 GAAAGCTGTTCACATCAGAAATGAAATTTCTTAGCCAGAAAGAAAT...C 873
|||||.....:|||||.....:|||||.....:|||||.....:
366 sArgLeuLeuGlnValGlyMetLysHisLeuGlnProGluAsnLeuMetA 383
874 AAGAGCTGAAGACAGACATAGACTTCATGCTGACTTCGACGAAAACTGA 923
:|||||.....:.....:.....:.....:.....:.....:
383 rGSerLeuLysProLeuSerGlyTyrMetArgGluLeu.....:Thr 396
924 AAGCAGAAAGATGAAACAGACACCTCACAGAGTGGGAGAGACGAGCTG 973
:|||||.....:|||||.....:|||||.....:|||||.....:
397 AlaArgGluLeuGlyPheGluLeuSerSerAsnAsp.....: 408
974 CTGGCAGACAGACAGACAGACTCTGGAGAT.....:G 1005
:|||||.....:|||||.....:|||||.....:|||||.....:
409 .metSerArgGluAsnAsnAspSerAlaAspGluLysAlaSerThrGlyA 425
1006 CGAAGACACACCCCTTCAAGATGCCAGCAAGAAATAAAGAAAGAACG 1055
|||||.....:|||||.....:|||||.....:|||||.....:
425 lAgLluGluProArgThrProAlaGlnLeuGluGluGluLysArg 441
1056 AAGATATGCTTGAAGATTAAGTTTAAAGC.....:AA 1087
|||||.....:|||||.....:|||||.....:|||||.....:
442 ThrLeuArgLeuGlnAsnAlaProLysThrHisPheGluProTyrValAs 458
1088 TAATTTAGACAGACAGCAG.....:GAGCAGCTTGACCGCATCGGAAG 1131
:|||||.....:|||||.....:|||||.....:|||||.....:
458 pAspLeuThrGlyGluGluLleLeuAlaGlnValAlaGlnAsnAlaAlaGln 475
1132 AATCTGGAGAGAAAGCTGACACGCGGCTGTGAACAGTACGTGCGAATT 1181
:|||||.....:|||||.....:|||||.....:|||||.....:
475 lnaLysGlnArgValAsnAlaLysLysAsnAlaHisLysMetGlyPhe 491
1182 GAAAGAGAAAGAACAGATGAACAACAGAGCTTGGACATTCACATCTCT 1231
|||||.....:|||||.....:|||||.....:|||||.....:
492 LeuArgGlnMetLysAspGlyThrThrLeuAsnLeuValLleLysG 508
1232 GCAATCCCATTTGAGAGACCCAGCTACTGCTGTGAGACACTGGGAATGA 1281
:|||||.....:|||||.....:|||||.....:|||||.....:
508 uGlu.....:AsnGluGlyProGluArgValAlaThrLleGlyAspL 522
1282 CAACGTGAGAGACTTCAGTCTGGAGTGAATATCTTGAGAGGGTTCAAGAG 1331
:|||||.....:|||||.....:|||||.....:|||||.....:
522 euValGlyLysLeuGlnLysGlySerAlaGlnLeuGlnThrArgGlnVal 538
1332 GATTAAGAGACAAAGTCACTCCAGTGTATATTTGAATTAATGAGGCCCTTA 1381
|||||.....:|||||.....:|||||.....:|||||.....:
539 AspLysArgAsnAlaValArgThrValLysSerLeuAsnTyrGlyAlaPh 555

```

```

1382 CAGTCTTATGACACCGATTATGACTGACATTTGCAATAATCAGCAGG 1431
      ::::::::::::::::::::
555 ealaSerPhalaProthRheasSerArgPheSerThleuSerIaG 572
      ::::::::::::::::::::
1432 ATGATTCGTATTAATCTATTATTCATGAGGAGACGTGATCTTCCA 1481
      ::::::::::::::::::::
572 luGluThrGlnLeuValLeuArgThrTyroIynAspAlaSerIaGlu 588
      ::::::::::::::::::::
1482 AGTGAATTCACATCCATGAGTTTGGCCAGCGCCAGATTAATCCGTA 1531
      ::::::::::::::::::::
589 TyraIaGluSerIleuGlnPheThrLysAsp...SerAsnTyroIyTh 604
      ::::::::::::::::::::
1532 TGTCATGCGCAGATGATTACTGATGTTTAAACAAAGGAGGCGATTCCA 1581
      ::::::::::::::::::::
604 rTrpIleAlaAsnClyleuLeuAspIleuThraAsnGlyGluHisSerI 621
      ::::::::::::::::::::
1582 GGACCCCTA.....CAA 1592
      ::::::::::::::::::::
621 ysSerIleuAspGluLeuTyraAsnMetGlnLeuHisSerTyroIyGlu 637
      ::::::::::::::::::::
1593 GAGATGAGATGTCATTCCTGGAAGATGAAGCCCATCTAGACACTTGA 1642
      ::::::::::::::::::::
638 GluIleGluLysCysPheGluGlnGlu.....G1 648
      ::::::::::::::::::::
1643 CACAGGAAAAAATGAGCAGATTACAGAAATAGACCCACGAGCGCTT 1692
      ::::::::::::::::::::
648 urThSerSerGlnGlnGluThrThraIaGlnIleGluGlnGluTyroI 665
      ::::::::::::::::::::
1693 TGGACTCCAGTACATCAGACAGGCTCATAGCGGTGAAGAGTAAACAA 1742
      ::::::::::::::::::::
665 ysTyroIynAsnThrHisValAspPheLysGlyLeuGlnSerLeuGly 681
      ::::::::::::::::::::
1743 TTTGGCGCTTCAGTTGAAGTTTGAAGTGAAGAGCTGAATAA.... 1787
      ::::::::::::::::::::
682 leuGlyIleAspValSerPheLeuAspGlyMetGluIaGluMetLys 698
      ::::::::::::::::::::
1788 .....TTCAGAAAGAACTTGATGAGACCCACCGAT 1818
      ::::::::::::::::::::
698 rTyroIyLeuAsnArgArgMetHisGlnHisLeuSerGlnAsnLeuTh 1715
      ::::::::::::::::::::
1819 TGGTCAGGAAATCCAGAAAGCCAGAAATGAGCTTGACACACGAGCC 1868
      ::::::::::::::::::::
715 euIleGlyLysLeuArgValAlaGlnHisAspArgLeuSer...Gln 730
      ::::::::::::::::::::
1869 CTTGGGAACATGATCTGCTCTGGGCTCCCTCATC...AGAGAATGA 1914
      ::::::::::::::::::::
731 leuProAsnHisLeuGlyLeuValGlnProAlaGlyGlnGluIleG 747
      ::::::::::::::::::::
1915 TCTTGTGCAACAGTGCACCAATATCTTAAGA..ATTGCACAGCAAGTA 1963
      ::::::::::::::::::::
747 urThraIaGlnGlnLeuThrGlnGlnIleSerAspLeuAlaLysLysLeu 764
      ::::::::::::::::::::
1964 CTCACAGTATATCTGTAAGCAGCAGTATGCAAAAGCAATGGGAGAT 2013
      ::::::::::::::::::::
764 roProSerAlaIleAlaAspProTyraIaLeuArgLysAlaMetGly 780
      ::::::::::::::::::::
2014 TCC.....ATTCCTTCCCGCCGTCATGAAACAACTTTGTGATTT 2054
      ::::::::::::::::::::
781 SerTyraIaGlyLeuProProArgProValSerProArgValGlnLe 797
      ::::::::::::::::::::
2055 GACAGAAACACGTAAGACCT 2076
      ::::::::::::::::::::
797 urProGluLeuGlnGlnPro 804
      ::::::::::::::::::::
seq_name: sp_human:Q9H8M2
seq_documentation_block:
ID Q9H8M2 PRELIMINARY; PRT; 501 AA.
AC Q9H8M2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

```

```

DE CDNA FLJ13441 FIS, CLONE PLACE1002775, WEAKLY SIMILAR TO
DE PERGRIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK023503; BAB14591.1;
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain.1.
DR PRINTS: PR00503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
SQ SEQUENCE 501 AA; 55664 MW; 5C68F53097BA073C CRC64;

```

```

alignment_scores:
  Quality: 684.00      Length: 473
  Ratio: 2.178         Gaps: 11
  Percent Similarity: 66.385      Percent Identity: 34.672

```

```

alignment_block:
US-09-687-230-1 x Q9H8M2 ..

```

```

Align seg 1/1 to: Q9H8M2 from: 1 to: 501

```

```

549 GAAGTGAAGACAGACACCCCTTCAAGAGCTTGAATCAACTGATGAGACA 598
      ||| ::::::::::::::::::::
19 GluAsnGluSerThrProIleGlnGlnLeuLeuGlnHisPheLeuArgL 35
      ::::::::::::::::::::
599 ATTGCAGAGAAAGATCCAAAGCTTTCTTTCATTTCCGTGACATGATT 648
      ::::::::::::::::::::
35 nleuGlnArgLysAspProHisGlyPhePheAlaPheProValThraAsp 52
      ::::::::::::::::::::
649 TTTATGCTCTGGCTACTCCATGATCATTTAAACACCAATGATTTTACT 698
      ::::::::::::::::::::
52 laIleAlaProGlyTyroSerMetIleIleLysHisProMetAspPheGly 68
      ::::::::::::::::::::
699 ACCATGAAAGAAAGATCAAGAACATGACTATCAGCTCAGCTAGAGACT 748
      ::::::::::::::::::::
69 ThrMetLysAspLysIleValAlaAsnGluTyroLysSerValThrGluPn 85
      ::::::::::::::::::::
749 AAAGATTAACCTCAACTAATGTGTAATATGCCATGATTTACAAATAAC 798
      ::::::::::::::::::::
85 elysAlaAspPheLysLeuMetCysAspAsnAlaMetThrTyraAsnArg 102
      ::::::::::::::::::::
799 CACAGACATTTATTATAAGCTGCAAGAGAGCTGTGACCTAGAGATG 848
      ::::::::::::::::::::
102 roAspThrValTyroTyroLysLeuAlaLysLysIleLeuHisAlaGlyPhe 118
      ::::::::::::::::::::
849 AAAATCTTATGCCAGAAAGATTCAGAGCTGAGAGAGAGCATGACTT 898
      ::::::::::::::::::::
119 LysMetCysSerLysGluArgLeuLeuAlaLeuLysArgSerMetSerPn 135
      ::::::::::::::::::::
899 CATGGCTGACTTCAGAAACATCGAAAGAGCAAGAAAGATGAAACAGACCT 948
      ::::::::::::::::::::
135 emetGlnAspMet.....AspPhe 142
      ::::::::::::::::::::
949 CACAGAGTGGGAGAGAGAGGCTGCTGCGCAGAGAGAGAGAGAGCTCT 998
      ::::::::::::::::::::
142 erGlnGlnAlaAlaLeuLeuGly.....AsnGluAspThr 153
      ::::::::::::::::::::
999 GGAGATGCCAGACACACAGCGCTTCAAGAGTCCC.....AG 1033
      ::::::::::::::::::::

```



```

154 AlaValGluGluProValProGluValValProValGlnValGluThrAl 170
1034 CAAAGAAATAAAGAAAGAAAGAAATATGCTTGAAGTAAGTTAA. 1082
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
170 alystysSerLysLysProSerArgGluValIleSerCysMetPheGluP 187
1083 .....AGCAATAAATTTACAGACAGACG 1103
      |||:|||||:|||||:|||||:|||||:|||||:
187 rogluGlyAsnAlaCysSerLeuThrAspSerThrAlaGluLuhIstVal 203
1104 CAGGAGCAGCTTGACCGCATGCTGAAAGAAATCGAGAAACATCAGC 1153
      |||:|||||:|||||:|||||:|||||:|||||:
204 LeuAlaLeuValGluHisAlaAlaAspGluAlaArgAspArgIleAsnAr 220
1154 GCGGCTTGAGACAGCTGACGTGCGAATTTGAAAGAAAGAAACAGATGNA 1203
      |||:|||||:|||||:|||||:|||||:|||||:
220 gPheLeuProGlyGlyLysMetGlyTyrLeuLysArgAsnGlyAspGlyS 237
1204 CACGACAGTGGGACTTCCATCTCTGTGATCCATGTGTAGAGACCA 1253
      |||:|||||:|||||:|||||:|||||:|||||:
237 eTLeuLeuTyrSerValValAsnThrAlaGluProAspAlaAspGluGlu 253
1254 GGCTACTGCTGCTGAGACTGGAGATGACACTGGAGACTCAGCTGG 1303
      |||:|||||:|||||:|||||:|||||:|||||:
254 GluThrHisProValAspLeuSerSerLeuSerSerLysLeuLeuProGl 270
1304 AGTGAATACCTTTCAGGGGTTCCAAAGAGATAAAGAAACAAAGTCACTC 1353
      |||:|||||:|||||:|||||:|||||:|||||:
270 yPheThrThrLeu...GlyPheLysAspGluThrArgAsnLysValThr. 285
1354 CAGGTGTATTGATTTGATGGCCCTACAGTCTTATGACCCGATTTAT 1403
      |||:|||||:|||||:|||||:|||||:|||||:
286 .....PheLeuSer...SerAlaThrThrAlaLeuSerMetGlnsN 298
1404 GACTCCACATTTGCAATATATGCAAGAGATGATTCGATTTATATATTC 1453
      |||:|||||:|||||:|||||:|||||:|||||:
299 AsnSerValPheGlyAspLeuLysSerAspGluMetGluLeuLeuTyrSe 315
1454 AACCTATGGGGAAGACTGTGATCTTCCAAAGTATTCAGCATCATGAGT 1503
      |||:|||||:|||||:|||||:|||||:|||||:
315 rAlaTyrGlyAspGluThrGlyValGlnLacSAlaLeuSerLeuGlnGluP 332
1504 TTGTGGCCACGTCGCAAGATTTATCGTATGTCATGGCAGATGCTTACG 1553
      |||:|||||:|||||:|||||:|||||:|||||:
332 heValLysAspAlaGlySerTyrSerLysValValAspAspLeuLeu 348
1554 GATGTTTAAACAAAGAGGGGATTCAGACCTTCAAGAGATGAG.. 1601
      |||:|||||:|||||:|||||:|||||:|||||:
349 AspGlnIleThrClyGlyAspHisSerArgThrLeuPheGlnLeuLysG 365
1602 .....ATGTCATTGCTGAGAGATGAGGCCCTACTAGCA 1635
      |||:|||||:|||||:|||||:|||||:|||||:
365 narGArgAsnValProMetLysProProAspGluAlaLysValGlyAsp 382
1636 CACTT...GACACAGAAAGAAATGAGACAGATTACAGAAATAGAGCCA 1682
      |||:|||||:|||||:|||||:|||||:|||||:
382 hTLeuGlyAspSerSerSerSerValLeuGluPheMetSerMetLys... 397
1683 CCAGGGCTTTGGACTCCAGTACTCAAGACAGGCTCATAGCGCTGAAAC 1732
      |||:|||||:|||||:|||||:|||||:|||||:
398 .....SerTyrProAspValSerValAspIleSerMet 408
1733 ACTAACAAATTTGGCGTTCAGTTGAAAGTTTTCAGCTCTGAGAAAGCTG 1782
      |||:|||||:|||||:|||||:|||||:|||||:
408 tLeuSerSerLeuGlyLysValLysLysGluLeuAspProAspAspSerH 425
1783 AAATATTCAGAGAAACTGATGAGACCAACAGATGCTCAGGGAGACCTC 1832
      |||:|||||:|||||:|||||:|||||:|||||:
425 tSLeu.....AsnLeuAspGluThrThrLysLeuLeuGlnAspLeu 438
1833 CAGGAAGCCACAGATGACGTTTGACACACAGACCCCTGGAGACATGAT 1882
      |||:|||||:|||||:|||||:|||||:|||||:
439 HtSGluAlaGlnAlaGluArgGlyLysArgProSerSerAsnLeuSe 455

```

```

1883 CTGCTCTTGGGTCCTCA 1901
      |||:|||||:
455 rSerLeuSerAsnAlaSer 461
seq_name: sp_rdent:09CT78
seq_documentation_block:
ID 09CT78 PRELIMINARY; PRT; 174 AA.
AC 09CT78;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 18 DAYS EMBRYO CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:1190001E05, FULL INSERT SEQUENCE (FRAGMENT).
GN BRD7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleishmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitlaker C., Wilming L.,
RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK04429; BAB23299.1; -.
DR MGD; MGI:1349766; Brd7.
FT NON_TER 1
FT 1
SO SEQUENCE 174 AA; 19027 MW; CF47FE3B9D719D03 CRC64;
Alignment_scores:
Quality: 635.50 Length: 175
Ratio: 3.972 Gaps: 1
Percent Similarity: 91.429 Percent Identity: 75.429
Alignment_block:
US-09-687-230-1 x 09CT78 ..
Align seg 1/1 to: 09CT78 from: 1 to: 174
1593 GAGATGGAGATGCTCATTCGCGAAGATGAGAGCCCTACTAGACACTGGA 1642
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
1 AspLeuAspMetSerSerProGlnAspGluGlyGlnThrArgAlaLeuAs 17
1643 CACAGAAAGAAAGAAATGAGACAGATTACAGAAATAGAGCCAGGCGCTT 1692
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 pPheAlaLysGluAlaGlu...IleThrGlnIleGluThrProThrGlyArgL 33
1693 TGGATCTCAGTACTCAAGACAGGCTCATAGCGCTGAAAGCGTAAACAAT 1742
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
33 euGluSerSerSerGlnAspArgLeuThrAlaLeuGlnAlaValAlaThrThr 49
1743 TTGTCGTTCCAGTTCGAGTTTTCGCTGAGAAAGCTGGAATATTCGCA 1792

```



```

879 CTGAAGCAGATAGACTTC..... 899
900 .....ATGCTGACTTCAGAAAATCGA.....AGCAGA 930
272 ylleatgproleuAlaProValProLysGluArgThrMetAsnLysArgL 289
931 AAGATGGAACAGACACCTCAGAGAGTGGAGAGAGAGCTGCTGGCAG 980
289 yAlaValValLysAspGlyMetThrSerGluAsp.....CysLeuGln 303
981 .....AGAGAGAGAGAGACTCTGAGATGCCGAAGC 1012
304 ValAlaAspProLysValArgGluArgLeuSerAlaLysLeuProGluAl 320
1013 ACACGCCCTTCAGAGTCCAGCAAAAGAAATMAAAAGAAAGCAAGATA 1062
320 a.....AsnAsnProLysAsnLysYSM 328
1063 TCGTTGAAGATAAGTTTAAAGCAATAATTAGAGAGAGAGAGAGAG 1112
328 etelLysLeuGlyPheLeuSer..... 335
1113 CTTGACCCGATCGTGAAGAAATCTGGAGAAACCTGACAGCGCGCTGT 1162
336 .....GluLysAspGlyThrValValLeuAsnValVa 346
1163 GAACACTGACTGCGAA.....TTGAAAAGAAAAGAAACGATGGA 1203
346 lAlaGlyAspSerGluAspLysLeuGluAsnAsnAlaProArgArgV 363
1204 CAACGAGCTGGGACTTCCTCATTCTGGATCCATTGAGAGAGAGCA 1253
363 alhrltleGlyAspIleValGlyProLeuGluLuu.....GlyThrPro 377
1254 GGGTACTGCGTGTGAGACTGGGAATGACACTGGAAAGCTTACGCTGG 1303
378 Gly.....MetIleGlnMetAlaAspHisArgLeuPheSer... 389
1304 AGTGAATACTTTCGAGGGGTTCAAGAGATAAAGAACAAAGTCACTC 1353
390 .....GlnAlaLeu 392
1354 CAGTGTATATTGATATTATGGCCCTACAGTTCCTATGACCCGATTAT 1403
392 roValAsnTyrlLeuAsnTyrlGlyProTyrSerSerPheAlaProMetTyr 408
1404 GACTCCCACTTTGCCAATATACAGCAAGAGATTCGTGATTATCTATTG 1453
409 AspSerThrTrpAlaThrMetThrLysGluAspThrAspLeuPheLeuAr 425
1454 AACCTATGGGGAAGACTGTGATCTTCACAGTGATTCAGCATCAGTAGT 1503
425 gThrlTyrlGlyAspLysSerAsnAlaSerPValMetSerMetArgArgp 442
1504 TTTTGGCCAGCTGCCAAGATTATCCGATATGTCATGGACGATAGTTACTG 1553
442 heValGlyAspCysProGluPheSerGluIleIleGly...SerLeuLeu 457
1554 GATGTTTAAACAAAAGAGCGGCTTCAGAGACCTTACAGAGATGAGAT 1603
458 AspThrLeuThrAspGlyGluHisSerLysThrMetLysGluLeuGlu.. 473
1604 GTCATTTGCTGAGAGATGAAGGCCATACTAGACACTTGACACAGGAAGA 1653
474 .....AsnAlaGlyLysG 478
1654 AAATGGAGCAATTAAGAAATAGAGCCACAGGCGTTGGACTCCAGT 1703
478 lValValLysGlu.....GluValAspAsnAspGlyTrpLysAsnGluThr 492
1704 ACTCAAGACAGGCTCATAGCGCTGAAGACACTAAACAATTTGGCGTTCC 1753

```

```

493 ValLeuSerLeuIleAspAspValSerSerIleSerAsnLeuGlyIleGl 509
1754 AGTGAAGCTTTTACCTCTGAGAAAGCTGAATATTC..... 1790
509 uThrGlyPheLeuAsnAspIleArgGlnIleValLeuValProAlaValG 526
1790 ..... 1790
526 lUserAsnIleGluAsnAsnIleProGluPheMetAsnGluValAsnHis 542
1791 .....CAGAAAGAACTGATGAGACACCGAGATTCCTCAGGAGACT 1831
543 MetAsnValGlnGlnGlnLeuAsnHisSerLysGlnLysValLysAspLe 559
1832 CCAGAGACCCAGCAATGAACTTTCAGACACGACCCCTCGGGAACATGA 1881
559 uAlaHisIleGlnGlnHisArgLeuValGlnGlnProProMetIleM 576
1882 TCTGCTCTTGGGCTCCCTCATTACAG..... 1906
576 etSerValGlnGlyValGlyGlnIleGlnGlnLysLeuAlaGluAsnLeu 592
1907 GAATGCAATCTTTCGAGACAGTACCAATAT 1939
593 GlnGlnHisLeuAlaHisGlnMetThrThrHis 603

```

```

seq_name: sp_human:Q9H7R9
seq_documentation_block:
ID Q9H7R9 PRELIMINARY; PRT; 233 AA.
AC Q9H7R9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CDNA FLJ14330 FIS, CLONE PLACE400261, WEAKLY SIMILAR TO
DE PERGRIN.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Tojiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RA "NEDO human cDNA sequencing project."
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024392; BAB14907.1; -
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR SEQUENCE 233 AA; 26233 MW; FBF96245BD88F4EB CRC64;

```

```

alignment_scores:
  Quality: 391.00 Length: 106
  Ratio: 4.116 Gaps: 0
  Percent Similarity: 89.623 Percent Identity: 65.094

```

```

alignment_block:
US-09-687-230-1 x Q9H7R9 ..
Align seg 1/1 to: Q9H7R9 from: 1 to: 233

```

```

549 GAAGTGAAGACAGACCCCTCAAGAGCTTGAATCACTAGTACAGACA 598
19 GluAsnGlnSerThrProIleGlnGlnLeuLeuAspHisPheLeuArgL 35

```

599 ATGACAGAAAGAAATCCAGTCTTTCTTTCTTCTTCTGACGATT 648
 35 nleuglnatrglyasprhnsiglypnehealaphneprovalthrpsa 52
 649 TTATGCTCTGCTACTCATGATCATTTAAACCAACCAATGATTTAGT 658
 52 lailealaprogllyrsermetlellelyshsiprometasphnegly 68
 699 ACCATGAAGAAAGATCAAGAACATGACTATCATGTCATGAAAGACT 748
 69 ThmetelysaspysilevalalaansglutrylyserValthrcluph 85
 749 AAGAGTAAGTCAACTAATGCTACTAATGCGATGATTACATTAAC 798
 85 elysalalaspheylseumetccysaspsnalamethrlyrasnarp 102
 799 CAGACAGCATTTATTAAGCTGCAAGAGAGCTGTTCACGACGAAAG 848
 102 roaspthrVallyrlyrlyslualalyelyleuhsialaglyphe 118
 849 AAATCTTAGCCAGCA 866
 119 lysmetleuserlysgln 124

seq_name: sp_human:Q9H505

seq_documentation_block:

ID Q9H505 PRELIMINARY; PRT; 362 AA.
 AC Q9H505.
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-DEC-2001 (TrEMBLrel. 16, Last sequence update)
 DE CDNA: FLJ23177 FIS, CLONE LNC10649.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Oho T.,
 RA Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,
 RA Odayashi M., Nishii T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isonogi T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026830; BAB15565.1;
 SO SEQUENCE 362 AA; 39411 MW; 3B8240D7C70E44A5 CRC64;

alignment_scores:

Quality: 274.00 Length: 307
 Ratio: 1.420 Gaps: 8
 Percent Similarity: 62.866 Percent Identity: 25.733

alignment_block:

US-09-687-230-1 x Q9H505 ..

Align seg 1/1 to: Q9H505 from: 1 to: 362

1032 AGCAAAAGAAATTAAGAAAGAAAGAAAGATGCTTGAAGATAAGTTAA 1081
 31 Alalulysrlyserlyslspsrseratrgluvalilesercysmetpnehl 47
 1082 A.....AGCAATTAATTGAGAGAG 1101
 47 uproglulylasnalacysserleuthraspserrthralagluglulhiv 64
 1102 AGCAGAGACGCTTGACCGCATGCGTGAAGAAATCTGAGGAGAACTGACG 1151
 64 alleualaleuvalglulhialalalaspriulalargarparglilean 80

1152 AGCGGCTTGTAACAGTCAGTCCGCAATTTGAAGAGAAACAGATGG 1201
 81 Argpneleuproglyglylsmetgllyrlyreuleysarargansglyaspgl 97
 1202 AACACGAGCGTGGGACTTCCATCTGTCGATGTCGATGTCGATGTCGATG 1251
 97 yserleuleuylrserValvalasnthralaglukspslaspglug 114
 1252 CAGGCTACTGCCGTGAGACTGGGAATGACACTGGAAGACTTCAGTCT 1301
 114 lugluthrhisprovalaspserleuserSerleuserSerleuserPro 130
 1302 GGAGTGAATGATTTGACAGGGTTCAGAGATTAAGAGAAAGAAAGTAC 1351
 131 Glypnehrthrleu...GlypnehsaspslualargaranslyValth 146
 1352 TCACAGTGTATTAATTTGATTAATGAGCCCTACAGTCTTATGACCGGAT 1401
 146 r.....pneuser...SerAlaThrThrAlaLeusermetcIna 159
 1402 ATGACTCCACATTTGCAATATTCAGCAAGGATGATTTGATTAATCTAT 1451
 159 snasrserValpneglyaspneulysseraspglumetgluleuleutyr 175
 1452 TCAACCTAAGGGAAGACTGATCTTCCAGATGATTTGACGACCTCA 1501
 176 SerAlatyrglyaspgluthrlyValglncysAlaleuserleuIngl 192
 1502 GTTTTGGCCAGCGCCAGATTAATTCGATGTCATGCGATGATGATTTAC 1551
 192 upheVallyaspAlaaglyserlyserlyslsValvalaspneul 209
 1552 TCGATGTTTAAACAAAGAGAGGATTCGAGACCCCTACAGAGATGGAG 1601
 209 eunaspglulthrlyglyaspshiserarqthrleupneuglneulys 225
 1602ATGTCATGCTGAGATGAGAGGCACTACTAG 1633
 226 GlnarargansvalPrometllyspProaspplualalalyValalyas 242
 1634 GACACT...GACACAGAGAAAGAAATGAGCAGATTTACAGAGATGAGC 1680
 242 pthrleuglyaspserSerSerSerlyValleuglupneuserMetlys. 258
 1681 CACCAGGCGTTGACACTCAGACTCAGACAGCGCTCATAGCGCTGAA 1730
 259SetTyrProaspValSerValaspIleSer 268
 1731 GCAGTACAAATTTGGCGTCCAGTTGAAGTTTGGACTGGAAGAGC 1780
 269 MetleuserSerleuInglysVallyslsgluLeuaspProaspaspe 285
 1781 TGAATATTCAGAGAAAGAACTGATGAGACCAACGATTTGCTCAGAGAAC 1830
 285 rHisleu.....AsnleuaspgluthrlyslsleuInglaspI 299
 1831 TCCAGAGACCCAGATGACGTTTGAAGCAGCAGACCCCTGGAAGCATG 1880
 299 euhlsleuInglInglInglInglInglInglInglInglInglInglIngl 315
 1881 ATCTGCTCTTGGTCCCTCA 1901
 316 SerSerleuserSerAlaSer 322

seq_name: sp_human:Q9Y403

seq_documentation_block:

ID Q9Y403 PRELIMINARY; PRT; 715 AA.
 AC Q9Y403.
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

"Sequences cloned from human fetal thymus cDNA library."
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF167351; AAD4658.1; -
 SQ SEQUENCE 56 AA; 6396 MW; 431E06FC3D9D7430 CRC64;

alignment_scores:

Quality: 252.00 Length: 55
 Ratio: 4.846 Gaps: 0
 Percent Similarity: 94.545 Percent Identity: 85.455

alignment_block:

US-09-687-230-1 x Q9SN19 ..

Align seg 1/1 to: Q9UN01 from: 1 to: 56

1266 GTGAGCTGTGGAATGACACTGGAGACTTCAGCTGAGTACTTT 1315
 1 MetArgLeuGlyMetThrGlyArgLeuGlnSerGlyValAsnIleLe 17
 1316 GCAGCGCTTCAAGAGCATAAAGAACAAAGTCACCTCAGCTTTATAT 1365
 17 uGlnGlyPheLysGlnAspLysArgAsnLysValThrProGlyLeuTyrL 34
 1366 TGAATTATGGGCCCTACAGTTCTTATGACCGCATTTGACTCCACATT 1415
 34 euAsnTyrGlyProIlePheSerSerTyrValProHisTyrAspSerThrPhe 50
 1416 GCAAAATTCAGCAAG 1430
 51 MetAsnIleLysGln 55

seq_name: sp_plant:Q9SN19

seq_documentation_block:

ID Q9SN19 PRELIMINARY; PRT: 952 AA.
 AC Q9SN19:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE SIMILAR TO DJ522J7.2.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE.
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 3, PAC
 RT clone: P0043E01."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF000615; BAA85417.1; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 SO SEQUENCE 952 AA; 105312 MW; 91D8E06A5A1769D CRC64;

alignment_scores:

Quality: 244.50 Length: 802
 Ratio: 0.745 Gaps: 40
 Percent Similarity: 40.898 Percent Identity: 22.195

alignment_block:

US-09-687-230-1 x Q9SN19 ..

Align seg 1/1 to: Q9SN19 from: 1 to: 952

42 CCTGGGGGGGGGCTCTCGGGGGCCGCCGCTCGGCTGGCCTG 91

||||| ||| ::::: ||| |||||
 20 ProArgHisAlaValAspArgSerHisThrSerProProPro..... 33
 92 GCCCGACCCGGAACCGGGCCGACGCGCTCGGCTCGGCGC..... 132
 34ProProAsnAlaProAlaArgGlyArgProIle.ArgAsnArgLeu 47
 133 ...GGGGGGCGGACCGGGGCCCGGTGGAACATGGGCAAGAACAGCAAG 179
 48 HisGlyAlaLeuGlnProAsnProAlaSerMetThrAlaProArgGly 64
 180 AAGCAAGAATGGGCAACAAACACCTCTACGAGAGATGTAGAGAACCCCT 229
 64 sAsnThrLysArgThrGlnSerHisThrHis.....Pro. 75
 230 GAACCTGCTCCTCAAGTAGAGGAGGAAGTAGTACCGACCTTCAC.. 277
 76 AspProPheProAlaLysAlaThrProArgProHisHisHisHisHis 92
 277 277
 92 sLeuLeuLeuProProProProProArgProSerMetAlaLysThrArgL 109
 277 277
 109 ySaAlaAlaAlaProProProProProProProProAlaGluThr 125
 278GGGACCTCGGGGACGAGCTCCAGCTTCCTTCGAAG 312
 126 ProAlaArgArgLysGlyLysLysLysGlyArgProSerLeuLeuAspLe 142
 313 ACAAACGATCAGTACAAACAAAGACAGACAGAACCGNA.....AAG 356
 142 uGlnArgArgSerLeuArgLeuGlnAlaGlnAsnProSerProAlaProS 159
 357 AGAAAGAAAGAGAGAGACAGATCCAGGGGGAAGAAAG..... 395
 159 erProSerArgArgAspAlaAsn.ProSerAspGluAspAspGlyVal 175
 396GGGAGAAACGG...AGAAAGTTAAAGAG..... 422
 175 lGlySerGlyArgArgArgGlnLysArgLeuLysSerValLeuSerS 192
 422 422
 192 erSerGlyGlyGluArgPheAsnProPheHisArgArgGluGlnTyr 208
 422 422
 209 AlaAspLeuThrSerPheAspGlyPheAspLeuPheProProGlyLeuLe 225
 423GATMAAAGAACGAGATCGAGAC..... 446
 225 uCysCysLysArgCysProGlyLeuLysLysArgGlnLysSerSerPheP 242
 446 446
 242 heValCysGlnValIlePheSerPheAlaAlaLeuArgProGlySerArgLys 258
 447CGGCTGGAATGAGCGCAGAAAGATCTCCAGTG 481
 259 ArgLeuThrPheLeuArgPheHisGlyGlnPheAspSerAspGlyCysCys 275
 482 T.....CAGGCCCT.....GTGAGAT 498
 275 sSerMetLeuGlnAspAspGluAlaProAlaAlaAlaValValLysVal 292
 499 TAGACTGCTCCTCAGAAAGCTTCACAGACTCTTTAGCCAAACAGAA 548
 292 alGluValGlnGluLysLysLysValSerSerLysAlaThrGlyLys 308
 549GAACTAGAACAGACACCCCT..... 569

309 G1YA9LALALASerAspGlyLysProThrThrGlyThrProLeuProAs 325

570 .CAAGAAAGCTTTAAATCAACATGTATGACAGCAATTCGACAGAAAGATCCAA 618

325 ||| :|||:|||||

325 pLysLysLeuLeuLeuPheLeuAspArgLeuLysLysAspThrT 342

619 GTCCTTCTCTTTCATTTCCCTGTGACAGATTTATTTGTCCTCCGACATCC 668

342 :||| ||||| ||||| ||| |||

342 yrgLYalPheSerGluProValAspHisGluGluLeuProAspTrpHis 358

669 ATGATCATTTAAACCAACCAATGATTTTACTACCATGAAGAAAGAACATCGA 718

359 GlnIleIleLeuIHisProMetAspPheSerThrIleArgGluLysLeuLe 375

719 GAACAAATGATATACATCCATCATGAGAAAGATAAGTAAGTAACTCAATCA 768

375 |||||:|||||:|||||:|||||:|||||:|||||

375 uAsnAspSerTyrThrThrLeuGluLnPheLysAsnAspValPheLeuL 392

769 TGTGTACTAATGCCATGATTTACATTAACCAACAGACACATTTATTTATAA 818

392 eutHrSerAsnAlaMetSerTyrAsnSerAspAspThrValTyrTyrArg 408

819 GCTGCAAAAGAGCTGTGCATCAGGAATGAAGAAATTTAGCCAGGAAG 868

409 GlnLeu. ||||| |||||

869 AATTGACAGCTTGAAACAGACAGATAGACTTATGCTGACTTGCAAGAAA 918

417 gIleValSerValArgSerGly.....GlnLysA 427

919 CTCGAAGACGAAAGATGGAACAGACACTCTACAG..... 953

427 rglIleArgIleProPheGlySerAspAsnSerArgIleValArgIleArg 443

953 |||:|||||:|||||:|||||:|||||

444 IleArgIleArgIlePheSerProGlyTyrGluTyrGluTyrGlyPhePr 460

954 .AGTGGGAGAGACGGA.....GGCTGCTGGCAG..... 980

460 oSerAspThrAspAsnAsnSerAspIleArgLysAsnGlySerPheArgIleA 477

981 |||:|||||:|||||:|||||:|||||

477 rglValArgSerTyrLeuIleLeuIleArgIleArgAspAsnProTyrLeu 493

996 ...TTGGAAGATGCGAGACAGACAGCTTTCAGAGATCCCAAGCAAGAAA 1042

494 PheProTyrSerSerTyrProTyrSerPheSerTyrProValLysAsnMe 510

1043 TAA.....AAGAAGACAAAG 1059

510 tLysThrAsnMetAlaArgSerIleGluAlaLeuAlaLysLysAspPheG 527

1060 ATATGCTTGAGATTAAGTTTAAAGCAATTAATTTAGACAGACAGACAGCA 1109

527 LysnLeuArgGlnAla.....SerGluProGluGlu 538

1110 CAGCTTGACCGCATGCTGAAGAAATCTGAGAGAAAGCTGACACAGCGCT 1159

539 GlnGlnProLysThrValProArgArgGly..ArgProProLysTyrAla 554

1160 TGTGACAGATGCTGCGAATTGAAAGAAAGAAACCAATGACATGACCAACGA 1209

555LysLysIleGluLysThrGluAsnHis 563

1210 CGTTGGG.....ACTTCTCCATCTGTGGATCCC 1238

563 pAlSerProAspLeuSerAsnAlaLysThrLysSerAlaAspHisIaG 580

1239 ATTGTAGGAGACCGACGACTACGCTGTGTGAGATCGGAATGAAACAATCG 1288

580 LuthrIleArgLysArgLeuThrGlyAspArgThrAlaGlnAlaIle 596

```

1289 AAGACT.....TCAGCTCGAGTGAATC.....TTTG 1317
      ||| ||| ::: |||||
597 ThThArgAspSerProPheLeuGlnHisAsnThrProGlySerPheIle 613
      ||| ||| ::: |||||
1318 AGGGGTTCAAAGAGCATTAAGAAGACAAAGTCACCTCCAGTGTATATTGG 1367
      ||| ||| ::: |||||
      aglyLys.ArgThrAspArg..... 619
1368 AATTATGGGCCCTACAGTCTTATGCACCAGATTANGATCCACATTGCG 1417
      ::||| ||||| ::: :::|||
620 ...PheGlyAspTyrSerGlyProSerLysTyrGlyLysLysThrThrPr 635
      ::||| ||||| ::: :::|||
1418 AAATATGCAAGAGATGATGATTCGATTTAATCFATTAACCTAATGGGAG 1467
      ::||| ||||| ::: :::|||
635 cthrIleSerAspAspGluArg.....ArgSerThrTyr..... 646
      |||||
1468 ACTGTGATCTTCCAAAGTAGATTTCAGCATCATCATGATTTTGGCCACGTGC 1517
      ..... 646
646 ..... 646
1518 CAGATTATCCGATATGTCATGCGAGATAGTATTACTGATGTTTTAACAAA 1567
      ||| ||| :::||| ||||| :::|||
647 ...AspGlnGlnTyrPheHisSerSerProLeuPheSerAlaLeu.... 660
      |||||
1568 AGGAGGCGCATTCGAGCACCTCAACAAGACATGAGATGATTCCTCGTAG 1617
      ||||| ::: :::||| ::: :::|||
661 .GlyGlyGluArgLysValLeu....ValProValGlyLeuGlnGlnG 675
      |||||
1618 ATGAGAGCCATACATGAGACATT.....GAC 1643
      :::||| :::||| :::|||
675 InHisAlaTyrAlaArgSerLeuAlaArgPheAlaLysPheGlyPro 691
      |||||
1644 ACAGGAAGAAGATG....GAGCAGATTACAGAGTAGAGCCACCAG 1687
      :::||| :::||| :::||| :::|||
692 ValGlyTyrAspIleAlaLysArgIleArgArgLeuLeuProSerAl 708
      |||||
1688 GCGTTTGAGCTCCAGTACTCAAGACAGCGCTCATAGCGCTGAAGAAGCTGTA 1737
      |
708 Y.....T 709
seq_name: sp_plant:Q9SFX1
seq_documentation_block:
ID Q9SFX1 PRELIMINARY; PRT; 556 AA.
AC Q9SFX1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
RT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOHETICAL 62.1 KDA PROTEIN.
F15M4.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Uteback T.R.,
BA Barnstead M.E., Bowman C.L., White O., Nieman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC012394; AAF16663.1; -
DR InterPro: IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR SMART; PR00503; BROMODOMAIN.
DR PRINTS; SM00297; BROMO; 1.
```

DR PROSITE: PS50014: BROMODOMAIN_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 556 AA; 62060 MW; 03878B1E71C891C2 CRC64;

alignment_scores:

Quality: 236.00 Length: 447
 Ratio: 1.078 Gaps: 20
 Percent Similarity: 48.993 Percent Identity: 24.161

alignment_block:

US-09-687-230-1 x Q95FX1 ..

Align seg 1/1 to: Q95FX1 from: 1 to: 556

```

172 AGCACAAGAAAGCAGACGCGACAAACACCTTACGAGAGATATGTAGAG 221
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
51 SerSerArgAsnProAsnPhSerAsnArgSerAsnArgArgArgArgArg 67
   :: |||:: |||::: |||::|||::: |||::|||::: |||::|||:::
222 AAGCCCTTGAGCTGCTCCTCAAGATGAGGAGGAGAGAGATGACCGAGCT 271
   :: |||:: |||::: |||::|||::: |||::|||::: |||::|||:::
67 eRasnSerGlnAspAspAspArgGlnArgArgAspLysLysHisArgLeu 83
272 CTCACAGGCGACCTCGGGCGACGACTCCAGCCTCTTCGAAGCAAAAAGC 321
   |||::|||:: |||::|||::: |||::|||::: |||::|||:::
84 LeuHis.GlyLeuAsnSerHisGluGly..... 92
322 ATCATGACAAACACAGACAGACAAACCGAAAGAGAAAGAGAGAGAGAG 371
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
93 .....ArgAspSerSerAsnSerLysSerGlyGlyLysArg 104
372 AAGCAGATTCCAGGGGAGAGAGGAGAGAGAGAGAGAGAGAGATGAGAGA 421
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
105 Leu.....AspSerAspAlaArgAsnArgArgLysIle..... 115
422 GGATAAAAAGAGCGAGATCGAGACCGGGGTGAGAAATGAGCGAGAGAAAG 471
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
116 .....AspGlySerAspAsnThrGlyGlnLys. 124
472 AATCTCAGTGTACGCGCCCTGTGAGATTAGACTTCCTCTCGAGAGAGCT 521
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
125 .....AlaSerLysAla 128
522 CTCACAGCTCTTATAGCCAAACAGAGAGAGATGAGACAGACCCCTT. 569
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
129 ThrAspIleLeuGlnArgGlySerLeuValGlnSerThrProLeuPrt 145
570 .....CAAGACGCTTGATCAACGATGAGACATGTGACAGAGAAAGATC 615
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
145 cAspLysLysLeuLeuPhePheIleLeuAspArgValGlnLysLysAsp 162
616 CAAGTGTCTTCTTTCATTTCTGTGACTGATTTTATTTGCTCTGGCTAC 665
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
162 hrTyrglyValLysSerAspProAlaAspProGlnLeuLeuProAspTyr 178
666 TCATGATGATTAACACCCATGATTTAGTACATGAAAGAAAGAT 715
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
179 TyrIleIleIleLysAsnProMetAspPheThrLeuArgLysLysIle 195
716 CAAGACAAATGACTATCATGCTCATGAGAA.....CTAAGG 753
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
195 uGlnSerGlyValLysThrThrLeuGlnPheGlnAlaSerLeuGlnA 212
754 ATAACTTCAAACTAATGTACTAATGCCATGATTTACATTAACCCAGAG 803
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
212 spValPhe...LeuIleCysThrAsnAlaMetGlnTyrAsnSerAlaAsp 227
804 ACCATTATATATAAGCTGCAAGAGAGCTGTGCACTCAGGAATGAAT 853
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
228 ThrValTyrTyrArgGlnAlaArgAlaMetLeuGlnLeuAlaLysLys 243
854 TCTTAGCCAGAAAGAAATTCAGAGCTGAGAGCAGAGCATGACTTCATGG 903

```

```

243 ..... 243
904 CTGACTTCAGAAAACTCGAAAGCAGAAAGATGAGACAGACCTTCACAG 953
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
244 ..AspPheGlyAsnLeuArgGlnGlnSerAspGlyGlnGlnProValSer 259
954 AGTGGGAGAGACGAGGCTGCTGGCAGAGAGAGAGAGAGAGAGAGAGAG 1003
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
260 LeuSerGlnGlnProLysValValLysArgGlyArg..... 271
1004 TGCCGAGACACACGCTTCAAGAGTCCGACGAGAGAGAAATTAAGAAAG 1053
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
272 .....ProProGlnLysSerLysLeuLysLys. 279
1054 ACAAGATATGCTTGAAGATTAAGTTTAAACATTAATTTGAGAGAGAG 1103
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
280 .....GlnLeuGln 282
1104 CAGGAGCAGCTTGACCGCATGTAAGAGATCTGAGAGAAAGCTG..... 1148
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
283 GlnSerLeuIleAspArgThrThrSerAspIleSerAlaAspAlaAla 1299
1149 .....ACCAGCGGCTTGTGAACAGTCAGTGCAGAT 1179
299 aPheThrTyrAlaGlyAspSerSerArgLeuSerGlySer.....T 313
1180 TTGAAGAGAGAAACACAGATGAGACAGACGTTGGGAGCTTCCTCCTCT 1229
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
313 TyrAsnLeuArgLys.....AsnProSerTyrGlyPheArgHisAla 327
1230 GTGGATCCCATTTGAGAGACCGGCTACTGCTGCTGCTG..... 1268
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
328 GlnThrSerAlaValAsnLysTyrGlyMetLysAsnValAspLysAsn 344
1269 .....AGACTGGGATGACACTGAGACCTTCAG..TCTGAGAGTGA 1308
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
344 gArgAspThrTyrAsnGlnAsnSerAlaSerLeuGlnAspSerSerIleP 361
1309 ATACTTTCGAGGGGTTCAAGAGAGATTAAGAGAAAGTCACTCCAGTG 1358
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
361 heThrLeu.....LeuAspAspAsnLeuLysGlnLeuThrProVal 374
1359 ...TTATNTTGAATTAATGAGCCCTTCAGTTCTTATGACACCGCATTA 1405
   |||::|||::: |||::|||::: |||::|||::: |||::|||:::
375 GlyLeuLysAlaGlnTyrGlyTyrAlaArgSerLeuAlaArgTyrAla 391
1406 CTCC.....ACATTTGCAATATC 1424
   ::|||:: |||::: |||::|||::: |||::|||::: |||::|||:::
391 aAsnIleGlyProValAlaThrThrPheAlaAsnVal 403

```

THIS PAGE BLANK (USPTO)